

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2004, 02:14:44, Search time 4048 Seconds

(without alignments) 16296.467 Million cell updates/sec

Title: US-09-967-237-1

Perfect score: 1522

Sequence: 1 acagtcagccgcatgctcc.....tcaataataattataat 1522

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

```

1: gb_da:*
2: gb_hg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*
29: gb_ov:*
30: gb_ov:*
31: gb_ov:*
32: gb_ov:*
33: gb_ov:*
34: gb_ov:*
35: gb_ov:*
36: gb_ov:*
37: gb_ov:*
38: gb_ov:*
39: gb_ov:*
40: gb_ov:*
41: gb_ov:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1522	100.0	1522	AR074439	AR074439 Sequence
2	1522	100.0	1522	AR081119	AR081119 Sequence
3	1522	100.0	1522	AR085316	AR085316 Sequence
4	1522	100.0	1522	AR088064	AR088064 Sequence
5	1522	100.0	1522	AR104223	AR104223 Sequence
6	1522	100.0	1522	AR143487	AR143487 Sequence
7	1522	100.0	1522	AR171392	AR171392 Sequence
8	1522	100.0	1522	AR171563	AR171563 Sequence
9	1522	100.0	1522	BD243152	BD243152 MN gene a
10	1522	100.0	1522	AX330007	AX330007 Sequence
11	1522	100.0	1522	AX332607	AX332607 Sequence
12	1522	100.0	1522	AX333244	AX333244 Sequence
13	1522	100.0	1522	AX336174	AX336174 Sequence
14	1522	100.0	1522	HSNATUMN	X66839 H. sapiens M
15	1520.4	99.9	1639	BC014950	BC014950 Homo sapi
16	1517.4	99.7	1519	HAI10588	A010588 Homo sapi
17	1399	91.9	1399	AR095263	AR095263 Sequence
18	759.2	49.9	1665	MMU245857	A0245857 Mus muscu
19	457.8	30.1	1671	AB086322	A086322 Mus muscu
20	414.4	27.2	6521	AX795690	AX795690 Sequence
21	414.4	27.2	6521	AX822141	AX822141 Sequence
22	414.4	27.2	6521	AX825781	AX825781 Sequence
23	414.4	27.2	10898	AR074442	AR074442 Sequence
24	414.4	27.2	10898	AR081122	AR081122 Sequence
25	414.4	27.2	10898	AR085319	AR085319 Sequence
26	414.4	27.2	10898	AR088067	AR088067 Sequence
27	414.4	27.2	10898	AR104226	AR104226 Sequence
28	414.4	27.2	10898	AR143490	AR143490 Sequence
29	414.4	27.2	10898	AR171395	AR171395 Sequence
30	414.4	27.2	10898	AR171566	AR171566 Sequence
31	414.4	27.2	10898	BD243155	BD243155 MN gene a
32	414.4	27.2	187856	AL157874	AL157874 Human DNA
33	414.4	27.2	415	AF334629	AF334629 Homo sapi
34	413.4	27.2	415	AR074454	AR074454 Sequence
35	413.4	27.2	415	AR081134	AR081134 Sequence
36	413.4	27.2	415	AR085331	AR085331 Sequence
37	413.4	27.2	415	AR088079	AR088079 Sequence
38	413.4	27.2	415	AR104238	AR104238 Sequence
39	413.4	27.2	415	AR143502	AR143502 Sequence
40	413.4	27.2	445	AR171406	AR171406 Sequence
41	413.4	27.2	445	AR171577	AR171577 Sequence
42	413.4	25.0	445	BD243167	BD243167 MN gene a
43	380.2	25.0	10898	HSNMC9	Z54349 H. sapiens M
44	303.8	20.0	408	AX198178	AX198178 Sequence
45	303.8	20.0	521	AX197674	AX197674 Sequence

ALIGNMENTS

```

RESULT 1
LOCUS AR074439 1522 bp DNA 1linear PAT 28-AUG-2000
DEFINITION Sequence 1 from patent US 5955075.
ACCESSION AR074439
VERSION AR074439.1 GI:10001194
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1522)
AUTHORS Zavada,J., Pastorekova,S. and Pastorek,J.
TITLE Method of inhibiting tumor growth using antibodies to MN protein
JOURNAL Patent: US 5955075-A 1 21-SEP-1999;
FEATURES

```

source 1..1522  
/organism="unknown"  
/mol\_type="unassigned DNA"

Query Match 100.0%; Score 1522; DB 6; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 3,4e-311;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

```

QY 1 AAGTACGCGATGCTCCCTGTGCCCCCAGCCCTGCTCCCTGTGTATCCCGCC 60
DB 1 AAGTACGCGATGCTCCCTGTGCCCCCAGCCCTGCTCCCTGTGTATCCCGCC 60
QY 61 CCGCTCCAGGCGCTCACTGTGCACTGCTGTGCACTGCTGTGCTGTGCTGTGCT 120
DB 61 CCGCTCCAGGCGCTCACTGTGCACTGCTGTGCACTGCTGTGCTGTGCTGTGCT 120
QY 121 CCCAAGAGTTGCCCGGATGCAAGAGATTCCCGTTGGAGAGAGCTTTCTGGGAA 180
DB 121 CCCAAGAGTTGCCCGGATGCAAGAGATTCCCGTTGGAGAGAGCTTTCTGGGAA 180
QY 181 GATGACCACTGGGGGAGAGAGATCTGCCAGTGAAGAGATTACCCAGAGAGAGAT 240
DB 181 GATGACCACTGGGGGAGAGAGATCTGCCAGTGAAGAGATTACCCAGAGAGAGAT 240
QY 241 CCACCCGAGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTA 300
DB 241 CCACCCGAGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTA 300
QY 301 GAAGTTAAGCTTAATCAAGAAAGAGAGCTCCCTGAAGTTAGAGATCTA 360
DB 301 GAAGTTAAGCTTAATCAAGAAAGAGAGCTCCCTGAAGTTAGAGATCTA 360
QY 361 GAGGCTCTGAGAGATCTCAAGAAACCAGAAATATGCCACAGAGAGAGAGAGAT 420
DB 361 GAGGCTCTGAGAGATCTCAAGAAACCAGAAATATGCCACAGAGAGAGAGAT 420
QY 421 GACCAAGATCAATGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480
DB 421 GACCAAGATCAATGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480
QY 481 GGGGCGCGCTTCCAGATCCCGGTGATATCCCGGAGAGAGAGAGAGAGAGAG 540
DB 481 GGGGCGCGCTTCCAGATCCCGGTGATATCCCGGAGAGAGAGAGAGAGAGAG 540
QY 541 CTGCGCGCGCTTCCAGATCCCGGTGATATCCCGGAGAGAGAGAGAGAGAGAG 600
DB 541 CTGCGCGCGCTTCCAGATCCCGGTGATATCCCGGAGAGAGAGAGAGAGAG 600
QY 601 AACATGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 AACATGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 GGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 GGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 TCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 TCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 ACCGCTTTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 781 ACCGCTTTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 GCGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 GCGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 GAAAGAAATCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 GAAAGAAATCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

```

QY 961 CTGCCCTCTGACTTCAAGCCGCTACTTCCATATAGAGAGAGAGAGAGAGAGAG 1020  
DB 961 CTGCCCTCTGACTTCAAGCCGCTACTTCCATATAGAGAGAGAGAGAGAGAGAG 1020  
QY 1021 GCCAG 1080  
DB 1021 GCCAG 1080  
QY 1081 CACACCTCTCTGACACCCCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
DB 1081 CACACCTCTCTGACACCCCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 GCGAGCAG 1200  
DB 1141 GCGAGCAG 1200  
QY 1201 AGTCTCGAG 1260  
DB 1201 AGTCTCGAG 1260  
QY 1261 GCGCTGAGTTTGGGCTCTCTTTTGTCTGACAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1261 GCGCTGAGTTTGGGCTCTCTTTTGTCTGACAGAGAGAGAGAGAGAGAGAGAGAG 1320  
QY 1321 AGGAG 1380  
DB 1321 AGGAG 1380  
QY 1381 ACTGAG 1440  
DB 1381 ACTGAG 1440  
QY 1441 GAG 1500  
DB 1441 GAG 1500  
QY 1501 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1522  
DB 1501 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1522

RESULT 2  
AR081119  
LOCUS AR081119 1522 bp DNA linear PAT 31-AUG-2000  
DEFINITION Sequence 1 from patent US 5972353.  
ACCESSION AR081119  
VERSION AR081119.1 GI:10007847  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1522)  
AUTHORS Zavada,J., Pastorekova,S. and Pastorek,J.  
TITLE NM proteins, polypeptides, fusion proteins and fusion polypeptides  
JOURNAL Patent: US 5972353-A 1 26-OCT-1999;  
FEATURES  
source 1..1522  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 1522; DB 6; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 3,4e-311;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

```

QY 1 AAGTACGCGATGCTCCCTGTGCCCCCAGCCCTGCTCCCTGTGTATCCCGCC 60
DB 1 AAGTACGCGATGCTCCCTGTGCCCCCAGCCCTGCTCCCTGTGTATCCCGCC 60
QY 61 CCGCTCCAGGCGCTCACTGTGCACTGCTGTGCACTGCTGTGCTGTGCTGTGCT 120
DB 61 CCGCTCCAGGCGCTCACTGTGCACTGCTGTGCACTGCTGTGCTGTGCTGTGCT 120

```

121 CCCAGAGTGGCCCCGATGACAGAGATTCCCTTGGAGAGAGGCTCTTCTGGGGA 180  
 121 CCCAGAGTGGCCCCGATGACAGAGATTCCCTTGGAGAGAGGCTCTTCTGGGGA 180  
 181 GATGATCCCACTGGGCGAGAGAGATCTGGCCAGTGAAGAGATTCAACCAAGAGAGAT 240  
 181 GATGATCCCACTGGGCGAGAGAGATCTGGCCAGTGAAGAGATTCAACCAAGAGAGAT 240  
 241 CCACCCGGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACT 300  
 241 CCACCCGGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACT 300  
 301 GAATTAAGCCTTAATCAGAAAGAGAGAGGCTCCCTGAGATTAAGAGATCTACTCT 360  
 301 GAATTAAGCCTTAATCAGAAAGAGAGAGGCTCCCTGAGATTAAGAGATCTACTCT 360  
 361 GAGGCTCTGAGAGATCTCTCAAGAACCCCAAGATTAAGCCAGAGAGAGAGAGAT 420  
 361 GAGGCTCTGAGAGATCTCTCAAGAACCCCAAGATTAAGCCAGAGAGAGAGAGAT 420  
 421 GACCAAGTCAATTGGGCTATGAGAGGCAACCGCCCTGGCCCGGGTGTCCCAAGCTGC 480  
 421 GACCAAGTCAATTGGGCTATGAGAGGCAACCGCCCTGGCCCGGGTGTCCCAAGCTGC 480  
 481 GCGGGCCGCTTCCAGTCCCGGTGGAATATCCGCGCCCACTGCGGCTTCTGCGCGGC 540  
 481 GCGGGCCGCTTCCAGTCCCGGTGGAATATCCGCGCCCACTGCGGCTTCTGCGCGGC 540  
 541 CTGCGCCCTTGAATCTCTGGGCTTCAAGCTCCCGCGCTCCCAAGATCTGCGCTGC 600  
 541 CTGCGCCCTTGAATCTCTGGGCTTCAAGCTCCCGCGCTCCCAAGATCTGCGCTGC 600  
 601 AAATAATGCAAGTGTGCACTGACCTGCTCTTGGGCTTGAAGTGGCTTGGTTC 660  
 601 AAATAATGCAAGTGTGCACTGACCTGCTCTTGGGCTTGAAGTGGCTTGGTTC 660  
 661 GGGCGGAGTACCGGCTCTGCAAGCTGCACTGCGGCGCTCCCAAGATCTGCGCGGC 720  
 661 GGGCGGAGTACCGGCTCTGCAAGCTGCACTGCGGCGCTCCCAAGATCTGCGCGGC 720  
 721 TCGAGACACACTGTGAGAGGCAACCGCTTCCCTGCGAGATCCAGTGGTTCACCTGAC 780  
 721 TCGAGACACACTGTGAGAGGCAACCGCTTCCCTGCGAGATCCAGTGGTTCACCTGAC 780  
 781 ACCGCTTTCAGAGATGACAGAGGCTTGGGCGCCCGGAGAGGCTTGGCGCTTGGCC 840  
 781 ACCGCTTTCAGAGATGACAGAGGCTTGGGCGCCCGGAGAGGCTTGGCGCTTGGCC 840  
 841 GCTTTCTGAGAGAGGCGCGAAGAAAACAGTGCCTATGAGCAAGTTCCTGCTGCGCTTG 900  
 841 GCTTTCTGAGAGAGGCGCGAAGAAAACAGTGCCTATGAGCAAGTTCCTGCTGCGCTTG 900  
 901 GAAGAAATGCTGAGAGAGGCTCAGAGACTCAGAGCTCCAGAGCTGAGACATATCTGACCTC 960  
 901 GAAGAAATGCTGAGAGAGGCTCAGAGACTCAGAGCTCCAGAGCTGAGACATATCTGACCTC 960  
 961 CTGCGCTTGAAGTCAAGCGCTACTTCCAAATATGAGAGGCTCTGACTTACACCGCCCTGT 1020  
 961 CTGCGCTTGAAGTCAAGCGCTACTTCCAAATATGAGAGGCTCTGACTTACACCGCCCTGT 1020  
 1021 GCGCGAGGCTGATCTGAGACTGTGTTTAAACAGACAGATGCTGAGTGTCTGAGAGACTC 1080  
 1021 GCGCGAGGCTGATCTGAGACTGTGTTTAAACAGACAGATGCTGAGTGTCTGAGAGACTC 1080  
 1081 CACACCTCTCTGACACCTCTGAGGAGCTGTGAGCTGTGCTGAGTCAACGCTGAACTTCCGA 1140  
 1081 CACACCTCTCTGACACCTCTGAGGAGCTGTGAGCTGTGCTGAGTCAACGCTGAACTTCCGA 1140  
 1141 GCGAGGAGGCTTGAATGAGAGAGTGAATGAGGCTCTCTGCTGAGAGAGAGAGC 1200  
 1141 GCGAGGAGGCTTGAATGAGAGAGTGAATGAGGCTCTCTGCTGAGAGAGAGAGC 1200  
 1201 AGTCTCTGGGCTGCTGAGCAGTCCAGCTGAATTCCTGCTGCTGCTGAGTCAATCTTA 1260

1201 AGTCTCTGGGCTGCTGAGCAGTCCAGCTGAATTCCTGCTGCTGCTGAGTCAATCTTA 1260  
 1261 GCGCTGATTTTGGCTCTCTTTTGTGCTGTCAACAGAGTGGGCTTCTGTCAGATGAGA 1320  
 1261 GCGCTGATTTTGGCTCTCTTTTGTGCTGTCAACAGAGTGGGCTTCTGTCAGATGAGA 1320  
 1321 AGGACACACAGAGAGGGAACCAAGAGGCTGTGACTACCGCCAGCAGAGATAGCCGAG 1380  
 1321 AGGACACACAGAGAGGGAACCAAGAGGCTGTGACTACCGCCAGCAGAGATAGCCGAG 1380  
 1381 ACTGAGGCTTAAGGCTGATCTTGAAGATGTGAGAGAGCAGCAGAGAGATCTGAGG 1440  
 1381 ACTGAGGCTTAAGGCTGATCTTGAAGATGTGAGAGAGCAGCAGAGAGATCTGAGG 1440  
 1441 GAGGCGGATTAAGTCTGCTGCTGCTGCTCAATATGCACTTCTTTAACTGCAAGAAAT 1500  
 1441 GAGGCGGATTAAGTCTGCTGCTGCTGCTCAATATGCACTTCTTTAACTGCAAGAAAT 1500  
 1501 TTTAAATTAATTTAATTAAT 1522  
 1501 TTTAAATTAATTTAATTAAT 1522

RESULT 3  
 AR085316 1522 bp DNA linear PAT 01-SEP-2000  
 LOCUS Sequence 1 from patent US 5981711.  
 DEFINITION AR085316  
 ACCESSION AR085316  
 VERSION AR085316.1 GI:10012085  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1522)  
 AUTHORS Zavada, J., Pastorekova, S. and Pastorek, J.  
 TITLE MN-specific antibodies and hybridomas  
 JOURNAL Patent: US 5981711-A 1 09-NOV-1999;  
 FEATURES  
 source 1..1522  
 /organism="Unknown"  
 /mol\_type="unassigned DNA"

ORIGIN  
 Query Match 100.0%; Score 1522; DB 6; Length 1522;  
 Best Local Similarity 100.0%; Pred. No. 3,4e-311;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACAGTCACGCGCATGCTCCCTGTGCGCCAGCCCTGCTCCCTGTTGATCCCGGCC 60  
 1 ACAGTCACGCGCATGCTCCCTGTGCGCCAGCCCTGCTCCCTGTTGATCCCGGCC 60  
 61 CTGCTCCAGGCTCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 61 CTGCTCCAGGCTCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 121 CCCAGAGGTTGCCCGGAGTGAAGAGATTTCCCTTGGAGAGAGGCTTCTGAGGGA 180  
 121 CCCAGAGGTTGCCCGGAGTGAAGAGATTTCCCTTGGAGAGAGGCTTCTGAGGGA 180  
 181 GATGATCCCACTGGGCGAGAGATCTGCGCCAGTGAAGAGATTCAACCAAGAGAGAT 240  
 181 GATGATCCCACTGGGCGAGAGATCTGCGCCAGTGAAGAGATTCAACCAAGAGAGAT 240  
 241 CCACCCGGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACT 300  
 241 CCACCCGGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACT 300  
 301 GAATTAAGCCTTAATCAGAAAGAGAGGCTCCCTGAGATTAAGAGATCTACTCTGT 360  
 301 GAATTAAGCCTTAATCAGAAAGAGAGGCTCCCTGAGATTAAGAGATCTACTCTGT 360  
 361 GAGGCTCTGAGAGATCTCTCAAGAACCCCAAGATTAAGCCAGAGAGAGAGAT 420

Db 361 GAGGCTCTGGAGATCTCTCAAGAACCCCGAATATGCTCCACAGGGAACAAAGAGGGAT 420  
 Qy 421 GACCAAGATCATTTGGGCTATGGAGCGACCCGCTGAGCCCGGGTGTCCCAAGCTGC 480  
 Db 421 GACCAAGATCATTTGGGCTATGGAGCGACCCGCTGAGCCCGGGTGTCCCAAGCTGC 480  
 Qy 481 GCGGCGGCTTCCAGTCCCGGCTGATATCCGCGCCGAGCTGCGCGCTTGTGCGCGCC 540  
 Db 481 GCGGCGGCTTCCAGTCCCGGCTGATATCCGCGCCGAGCTGCGCGCTTGTGCGCGCC 540  
 Qy 541 CTGCGCCCTTGGAAATCTCTGGGCTTCCAGCTCCGCGCTCCCAAGACTGCGCTGCGC 600  
 Db 541 CTGCGCCCTTGGAAATCTCTGGGCTTCCAGCTCCGCGCTCCCAAGACTGCGCTGCGC 600  
 Qy 601 AACATGGCCACAGTGTGACCTGACCTCCGCTCTGAGGCTAGAGATGGCTCTGGTCCC 660  
 Db 601 AACATGGCCACAGTGTGACCTGACCTCCGCTCTGAGGCTAGAGATGGCTCTGGTCCC 660  
 Qy 661 GGGCGGAGTACCGGGCTCTGCAAGCTGCACTGCACTGGGGGCTGCAAGTCTGCGGGC 720  
 Db 661 GGGCGGAGTACCGGGCTCTGCAAGCTGCACTGCACTGGGGGCTGCAAGTCTGCGGGC 720  
 Qy 721 TCGGAGCACTGTGGAGAGGCCACCGTTTCCCTGCGGAGATCCAGTGTTCACCTGAGC 780  
 Db 721 TCGGAGCACTGTGGAGAGGCCACCGTTTCCCTGCGGAGATCCAGTGTTCACCTGAGC 780  
 Qy 781 ACCGCTTTGCGAGAGTTGACGAGGCTTGGGGGCGCCGCGGAGGCTGCGCTGTGTGCGC 840  
 Db 781 ACCGCTTTGCGAGAGTTGACGAGGCTTGGGGGCGCCGCGGAGGCTGCGCTGTGTGCGC 840  
 Qy 841 GCTTTCTGGAGAGAGGGCCCGGAAGAAACAGTGCCTATGAGCAGTTGCTGTCTGCTTG 900  
 Db 841 GCTTTCTGGAGAGAGGGCCCGGAAGAAACAGTGCCTATGAGCAGTTGCTGTCTGCTTG 900  
 Qy 901 GAAGAAATGCTGAGAGAGGCTCAGAGACTCAGGTCCAGGACTGAGCATATCTGACATC 960  
 Db 901 GAAGAAATGCTGAGAGAGGCTCAGAGACTCAGGTCCAGGACTGAGCATATCTGACATC 960  
 Qy 961 CTGCTCTGATCTTACGCGCTACTTCCAAATATGAGGGGCTCTGACTACAGCGCTGT 1020  
 Db 961 CTGCTCTGATCTTACGCGCTACTTCCAAATATGAGGGGCTCTGACTACAGCGCTGT 1020  
 Qy 1021 GCCCAGGCTGATCTGAGTGTGTTTAAACAGACAGTATGCTGAGTGTAGAGAGCTC 1080  
 Db 1021 GCCCAGGCTGATCTGAGTGTGTTTAAACAGACAGTATGCTGAGTGTAGAGAGCTC 1080  
 Qy 1081 CAGACCTCTCTGACACCTCTGAGGAGCTGTGTGACTCTCGGCTACAGCTGAATCTCGA 1140  
 Db 1081 CAGACCTCTCTGACACCTCTGAGGAGCTGTGTGACTCTCGGCTACAGCTGAATCTCGA 1140  
 Qy 1141 GCGAGCGACCTTGAATGAGGCGAGTGTGAGGCGCTTCCCTGCTGAGAGAGCAAC 1200  
 Db 1141 GCGAGCGACCTTGAATGAGGCGAGTGTGAGGCGCTTCCCTGCTGAGAGAGCAAC 1200  
 Qy 1201 AGTCTCGGCTGTGAGCGAGTCAAGCTGAATCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
 Db 1201 AGTCTCGGCTGTGAGCGAGTCAAGCTGAATCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
 Qy 1261 GCGCTGCTTGTGCTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
 Db 1261 GCGCTGCTTGTGCTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
 Qy 1321 AGGCAAGCAGAGAGGGAACCAAGGGGGTGTGAGCTACCGCCAGCAGAGGTAGCCGAG 1380  
 Db 1321 AGGCAAGCAGAGAGGGAACCAAGGGGGTGTGAGCTACCGCCAGCAGAGGTAGCCGAG 1380  
 Qy 1381 ACTGAGGCTTGAAGGCTGAGTCTTGGAGATGTGAGAGCAGCCAGAGGCACTGAGAGG 1440  
 Db 1381 ACTGAGGCTTGAAGGCTGAGTCTTGGAGATGTGAGAGCAGCCAGAGGCACTGAGAGG 1440  
 Qy 1441 GGAAGCGATACTGTCTGTCTGCTCATTTATGCACTTCTTTAACTGCCAAGAAATT 1500

Db 1441 GGAAGCGATACTGTCTGTCTGCTCATTTATGCACTTCTTTAACTGCCAAGAAATT 1500  
 Qy 1501 TTTTAAATTAATAATTATTAAT 1522  
 Db 1501 TTTTAAATTAATAATTATTAAT 1522  
 RESULT 4  
 AR088064 1522 bp DNA linear PAT 07-SEP-2000  
 LOCUS AR088064  
 DEFINITION Sequence 1 from patent US 5989838.  
 ACCESSION AR088064  
 VERSION AR088064.1 GI:10014827  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1522)  
 AUTHORS Zavada, J., Pastorekova, S. and Pastorek, J.  
 TITLE Immunological methods of detecting MN proteins and MN polypeptides  
 JOURNAL Patent: US 589838-A 1 23-NOV-1999;  
 FEATURES  
 source 1..1522  
 /organism="unknown"  
 /mol\_type="unassigned DNA"  
 ORIGIN  
 Query Match 100.0%; Score 1522; DB 6; Length 1522;  
 Best Local Similarity 100.0%; Pred. No. 3,4e-311;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ACAGTCAAGCCGATAGTCTCCCTGTGCGCCAGCGCCCTGCTGCTGTGATCCCGGCC 60  
 Db 1 ACAGTCAAGCCGATAGTCTCCCTGTGCGCCAGCGCCCTGCTGCTGTGATCCCGGCC 60  
 Qy 61 CCGCTCCAGGCTCACTGTGCACTGTGCTGTCACTGCTGCTTGTGATCGCTGCAT 120  
 Db 61 CCGCTCCAGGCTCACTGTGCACTGTGCTGTCACTGCTGCTTGTGATCGCTGCAT 120  
 Qy 121 CCCCAGAGTGTGCCCCGAGTGCAGAGAGATTCCTCCCTGGAGAGGCTCTTCTGGGGA 180  
 Db 121 CCCCAGAGTGTGCCCCGAGTGCAGAGAGATTCCTCCCTGGAGAGGCTCTTCTGGGGA 180  
 Qy 181 GATGACCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT 240  
 Db 181 GATGACCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT 240  
 Qy 241 CCACCCGAGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACT 300  
 Db 241 CCACCCGAGAGAGAGATCTACCTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 300  
 Qy 301 GAAGTTAAGCTTAATCAGAGAGAGGCTCCCTGAAGTTAGAGATCTACTACTGTT 360  
 Db 301 GAAGTTAAGCTTAATCAGAGAGAGGCTCCCTGAAGTTAGAGATCTACTACTGTT 360  
 Qy 361 GAGGCTCTGAGAGATCTCAAGAACCCCGAATATGCCCCAGGGAACAAAGAGGGAT 420  
 Db 361 GAGGCTCTGAGAGATCTCAAGAACCCCGAATATGCCCCAGGGAACAAAGAGGGAT 420  
 Qy 421 GACCAAGTCAATTTGGGCTATGAGAGCGACCCGCTGAGCCCGGGTGTCCCAAGCTGC 480  
 Db 421 GACCAAGTCAATTTGGGCTATGAGAGCGACCCGCTGAGCCCGGGTGTCCCAAGCTGC 480  
 Qy 481 GCGGCGGCTTCCAGTCCCGGCTGATATCCGCGCCGAGCTGCGCGCTTGTGCGCGCC 540  
 Db 481 GCGGCGGCTTCCAGTCCCGGCTGATATCCGCGCCGAGCTGCGCGCTTGTGCGCGCC 540  
 Qy 541 CTGCGCCCTTGGAACTCTTGGGCTTCCAGCTCCGCGCTCCCAAGACTGCGCTGCGC 600  
 Db 541 CTGCGCCCTTGGAACTCTTGGGCTTCCAGCTCCGCGCTCCCAAGACTGCGCTGCGC 600  
 Qy 601 AACATGGCCACAGTGTGACCTGACCTCCGCTCTGAGGCTAGAGATGGCTCTGGTCCC 660



Db 601 AACATGGCCACAGTGTGCACTGACCCCTCTCTGGGCTAGAGATGCTCTGGGCTCCC 660  
 Qy 661 GGGGGGAGTACCGGGCTCTGACGCTGACATCTGACATGGGGGGCTGACAGTGTGCGGAC 720  
 Db 661 GGGGGGAGTACCGGGCTCTGACGCTGACATCTGACATGGGGGGCTGACAGTGTGCGGAC 720  
 Qy 721 TCGAGACACTGTGAGAGGCAACCGTTTCCCTGCGAGATCCAGTGTGTCTGACCTGACG 780  
 Db 721 TCGAGACACTGTGAGAGGCAACCGTTTCCCTGCGAGATCCAGTGTGTCTGACCTGACG 780  
 Qy 781 ACCGCTTTTGCAGAGTTGACGAGGCTTGGGCGCCCGGAGGCTGCGCTGTGCGC 840  
 Db 781 ACCGCTTTTGCAGAGTTGACGAGGCTTGGGCGCCCGGAGGCTGCGCTGTGCGC 840  
 Qy 841 GCGCTTTTGAAGAGGGCCGGAAGAAACAGTGCCTATGACAGATGTGCTGCTGCTG 900  
 Db 841 GCGCTTTTGAAGAGGGCCGGAAGAAACAGTGCCTATGACAGATGTGCTGCTGCTGCTG 900  
 Qy 901 GAAGAAATCCCTGAGAGAGGCTCAGAGCTCAGGTCCAGAGCTGAGCATATCTGACCTC 960  
 Db 901 GAAGAAATCCCTGAGAGAGGCTCAGAGCTCAGGTCCAGAGCTGAGCATATCTGACCTC 960  
 Qy 961 CTGCGCTCTGACTTCAAGCGCTACTTCCATATGAGGGGTCTCTGACTTACACCGCTCTG 1020  
 Db 961 CTGCGCTCTGACTTCAAGCGCTACTTCCATATGAGGGGTCTCTGACTTACACCGCTCTG 1020  
 Qy 1021 GCCCAGGGTGTCACTGAGCTGTGTTTAAACAGACAGTGTGCTGAGTGTGAGAGCTC 1080  
 Db 1021 GCCCAGGGTGTCACTGAGCTGTGTTTAAACAGACAGTGTGCTGAGTGTGAGAGCTC 1080  
 Qy 1081 CACACCTCTCTGACACCCCTGAGGAGACCTGAGTGTGCTGAGTGTGAGTGTGAGTGTG 1140  
 Db 1081 CACACCTCTCTGACACCCCTGAGGAGACCTGAGTGTGCTGAGTGTGAGTGTGAGTGTG 1140  
 Qy 1141 GCGACGAGCGCTTGAATGAGGCGAGTGTGAGGCTCTCTGCTGCTGAGTGTGAGTGTG 1200  
 Db 1141 GCGACGAGCGCTTGAATGAGGCGAGTGTGAGGCTCTCTGCTGCTGAGTGTGAGTGTG 1200  
 Qy 1201 AGTCTCTGGGCTCTGAGGCACTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1260  
 Db 1201 AGTCTCTGGGCTCTGAGGCACTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1260  
 Qy 1261 GCGCTGCTTGTGGCT 1320  
 Db 1261 GCGCTGCTTGTGGCT 1320  
 Qy 1321 AGCAGACACAGAGGAGGAAACAAAGGGGTGTGAGCTACCGCCAGAGAGTGTGAGCAG 1380  
 Db 1321 AGCAGACACAGAGGAGGAAACAAAGGGGTGTGAGCTACCGCCAGAGAGTGTGAGCAG 1380  
 Qy 1381 ACTGAGAGCTTGAAGGCTGAGTGTGAGGAAATGAGAGGAGGAGGAGGAGTGTGAGG 1440  
 Db 1381 ACTGAGAGCTTGAAGGCTGAGTGTGAGGAAATGAGAGGAGGAGGAGGAGTGTGAGG 1440  
 Qy 1441 GAGAGCGGTACTGTCTGTCTGTCTGATTAAGCACTTCTCTTCTGAGCAAGAAAT 1500  
 Db 1441 GAGAGCGGTACTGTCTGTCTGTCTGATTAAGCACTTCTCTTCTGAGCAAGAAAT 1500  
 Qy 1501 TTTTAAATTAATTTATTAAT 1522  
 Db 1501 TTTTAAATTAATTTATTAAT 1522

RESULT 5  
 LOCUS AR104223 1522 bp DNA linear PAT 14-FEB-2001  
 DEFINITION Sequence 1 from patent US 6093548.  
 ACCESSION AR104223  
 VERSION AR104223.1 GI:12816931  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 Unclassified.

REFERENCE 1 (bases 1 to 1522)  
 AUTHORS Zavada, J., Pastorekova, S. and Pastorek, J.  
 TITLE Detection and quantitation of MN-specific antibodies  
 JOURNAL Patent: US 6093548-A 1 25-JUL-2000;  
 FEATUERS Location/Qualifiers  
 source 1..1522  
 /organism="unknown"  
 /mol\_type="unassigned DNA"

Query Match 100.0%; Score 1522; DB 6; Length 1522;  
 Best Local Similarity 100.0%; Pred. No. 3,4e-311;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAGTAGCGGATAGGCTCCCTGCGCCGAGCCCTTGGCTCTCTGTGATCCCGGC 60  
 Db 1 ACAGTAGCGGATAGGCTCCCTGCGCCGAGCCCTTGGCTCTCTGTGATCCCGGC 60  
 Qy 61 CTGCTCCAGGCTCACTGTGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 Db 61 CTGCTCCAGGCTCACTGTGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 Qy 121 CCCGAGGTTCCCGGATGAGAGGATTTCCCTTGGAGAGGCTCTTCTGGGAA 180  
 Db 121 CCCGAGGTTCCCGGATGAGAGGATTTCCCTTGGAGAGGCTCTTCTGGGAA 180  
 Qy 181 GATGACCCACTGGGCGAGAGGATCTGCGCAGTGAAGAGATTCACCCAGAGAGGAT 240  
 Db 181 GATGACCCACTGGGCGAGAGGATCTGCGCAGTGAAGAGATTCACCCAGAGAGGAT 240  
 Qy 241 CCAACCGGAGAGAGGATCTTACCTGAGAGAGGATCTTACCTGAGAGAGGATCTTACT 300  
 Db 241 CCAACCGGAGAGAGGATCTTACCTGAGAGAGGATCTTACCTGAGAGAGGATCTTACT 300  
 Qy 301 GAAGTTAAGCCTAATCAAGAAAGAGGCTCCCTGAAGTGAAGATCTTACCTGAT 360  
 Db 301 GAAGTTAAGCCTAATCAAGAAAGAGGCTCCCTGAAGTGAAGATCTTACCTGAT 360  
 Qy 361 GAGGCTCTGAGATCTTCAAGAACCCAGATTAATGCGCCAGAGAACAAAGAGGAT 420  
 Db 361 GAGGCTCTGAGATCTTCAAGAACCCAGATTAATGCGCCAGAGAACAAAGAGGAT 420  
 Qy 421 GACAGAGTCAATGGGCTGATGAGAGGACCCGCGCTGAGGCTGCGCTGCGCTGCG 480  
 Db 421 GACAGAGTCAATGGGCTGATGAGAGGACCCGCGCTGAGGCTGCGCTGCGCTGCG 480  
 Qy 481 GCGGCGCGCTTCAAGTCCCGAGTATCCGCGCCAGAGTCCGCGCTTCTGCGCGGC 540  
 Db 481 GCGGCGCGCTTCAAGTCCCGAGTATCCGCGCCAGAGTCCGCGCTTCTGCGCGGC 540  
 Qy 541 CTGCGCCCTTGAAGATCTCTGAGGCTTCCAGTCTCCGCGCTTCCAGAACTGCGCTGCG 600  
 Db 541 CTGCGCCCTTGAAGATCTCTGAGGCTTCCAGTCTCCGCGCTTCCAGAACTGCGCTGCG 600  
 Qy 601 AACCAATGGCAAGTGTCAACTGACCTGCTCTGAGGCTGAGAGTGTGAGTGTGAGT 660  
 Db 601 AACCAATGGCAAGTGTCAACTGACCTGCTCTGAGGCTGAGAGTGTGAGTGTGAGT 660  
 Qy 661 GGGCGGAGTACCGGGCTCTGAGTGTGATGATGAGGAGGCTGAGAGTGTGAGTGTGAG 720  
 Db 661 GGGCGGAGTACCGGGCTCTGAGTGTGATGATGAGGAGGCTGAGAGTGTGAGTGTGAG 720  
 Qy 721 TCGAGACACTGTGAGAGGCAACGTTTCCCTGCGAGATCCAGTGTGTTCACCTGAGC 780  
 Db 721 TCGAGACACTGTGAGAGGCAACGTTTCCCTGCGAGATCCAGTGTGTTCACCTGAGC 780  
 Qy 781 ACCGCTTTTGCAGAGTTGACGAGGCTTGGGAGGCTGCGGAGGCTGCGCTGTGCGC 840  
 Db 781 ACCGCTTTTGCAGAGTTGACGAGGCTTGGGAGGCTGCGGAGGCTGCGCTGTGCGC 840  
 Qy 841 GCGCTTCTGAGAGGAGGCGCGGAGAGAAACAGTGCCTATGAGAGTGTGCTGCTGCTG 900  
 Db 841 GCGCTTCTGAGAGGAGGCGCGGAGAGAAACAGTGCCTATGAGAGTGTGCTGCTGCTG 900

QY 901 GAAGAATCGCTGAGGAGGCTCAGAGACTCAGTCCAGAGCTGGAATATCTGCACTC 960  
DB 901 GAAGAATCGCTGAGGAGGCTCAGAGACTCAGTCCAGAGCTGGAATATCTGCACTC 960  
QY 961 CTGCTCTGACTTCAAGCCGCTACTTCCAAATATGAGGAGTCTGACTCAACGCGCTGT 1020  
DB 961 CTGCTCTGACTTCAAGCCGCTACTTCCAAATATGAGGAGTCTGACTCAACGCGCTGT 1020  
QY 1021 GCCCAGAGTGTCACTGAGCTGTGTTAACAGACAGTATGCTGAGTGTAAAGCACTC 1080  
DB 1021 GCCCAGAGTGTCACTGAGCTGTGTTAACAGACAGTATGCTGAGTGTAAAGCACTC 1080  
QY 1081 CACACCTCTCTGACACCCCTGTGAGGAGCTGAGTACTCTGAGCTACAGCTGAATTCGA 1140  
DB 1081 CACACCTCTCTGACACCCCTGTGAGGAGCTGAGTACTCTGAGCTACAGCTGAATTCGA 1140  
QY 1141 GCGACGAGCTTTGATGAGGAGTATGAGGCTGAGGCTGAGTCTGAGTGTAAAGCACTC 1200  
DB 1141 GCGACGAGCTTTGATGAGGAGTATGAGGCTGAGGCTGAGTCTGAGTGTAAAGCACTC 1200  
QY 1201 AGTCTCTGAGCTGCTGAGGCTGAGTATGAGTGTAAATTCCTGCTGAGTGTAAAGCACTC 1260  
DB 1201 AGTCTCTGAGCTGCTGAGGCTGAGTATGAGTGTAAATTCCTGCTGAGTGTAAAGCACTC 1260  
QY 1261 GCGCTGCTTTTGGCTCTCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
DB 1261 GCGCTGCTTTTGGCTCTCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
QY 1321 AGGACGACAGAGAGGAGCAAAAGGAGTGTGAGTATCCGCTGAGGAGTAAAGTAAAG 1380  
DB 1321 AGGACGACAGAGAGGAGCAAAAGGAGTGTGAGTATCCGCTGAGGAGTAAAGTAAAG 1380  
QY 1381 ACTGAGCTTAAAGTGTGATCTTGAAGATGTGAGAGCCAGGCAAGGATATGAGG 1440  
DB 1381 ACTGAGCTTAAAGTGTGATCTTGAAGATGTGAGAGCCAGGCAAGGATATGAGG 1440  
QY 1441 GAGCGGCTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
DB 1441 GAGCGGCTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
QY 1501 TTTTAAATTAATTTATTAAT 1522  
DB 1501 TTTTAAATTAATTTATTAAT 1522

RESULT 6  
AR143487  
LOCUS AR143487 1522 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 1 from patent US 6204370.  
ACCESSION AR143487  
VERSION AR143487.1 GI:15104773  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1522)  
AUTHORS Zavada,J., Pastorekova,S. and Pastorek,J.  
TITLES MN gene and protein  
JOURNAL Patent: US 6204370-A 1 20-MAR-2001;  
FEATURES  
source 1. 1522  
/organism="unknown"  
/mol\_type="unassigned DNA"

Query Match 100.0%; Score 1522; DB 6; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 3,4e-311;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTACGCCGATGCTCTCCCTGTGCCCCCAGCCCTGCTCTCTGTGATCCGAGC 60  
DB 1 ACAGTACGCCGATGCTCTCCCTGTGCCCCCAGCCCTGCTCTCTGTGATCCGAGC 60

QY 61 CCTGCTCAGAGGCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
DB 61 CCTGCTCAGAGGCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
QY 121 CCCAGAGGTTGCCCGGATGACAGAGGATTTCCCTTGGAGAGGCTTTCTGAGGAA 180  
DB 121 CCCAGAGGTTGCCCGGATGACAGAGGATTTCCCTTGGAGAGGCTTTCTGAGGAA 180  
QY 181 GATGACCACTGGGCGAGAGGATCTGCGCAGTGAAGAGGATTCACCCAGAGAGGAT 240  
DB 181 GATGACCACTGGGCGAGAGGATCTGCGCAGTGAAGAGGATTCACCCAGAGAGGAT 240  
QY 241 CCACCCGAGAGAGGATCTACCTGAGAGAGGATCTACCTGAGAGAGGATCTACCT 300  
DB 241 CCACCCGAGAGAGGATCTACCTGAGAGAGGATCTACCTGAGAGAGGATCTACCT 300  
QY 301 GAGTTAGCTTAATGAGAGAGGAGGCTCCCTGAGTGAAGTGAAGATCTACTGTT 360  
DB 301 GAGTTAGCTTAATGAGAGAGGAGGCTCCCTGAGTGAAGTGAAGATCTACTGTT 360  
QY 361 GAGGCTCCTGAGATCTCAAGAAACCCAGAAATATGCCACAGAGGACAAAGAGGAGAT 420  
DB 361 GAGGCTCCTGAGATCTCAAGAAACCCAGAAATATGCCACAGAGGACAAAGAGGAGAT 420  
QY 421 GACCAAGATCATTTGGGCTATGAGAGGAGACCCGCTGAGCCCGGCTGCTCCAGCTGC 480  
DB 421 GACCAAGATCATTTGGGCTATGAGAGGAGACCCGCTGAGCCCGGCTGCTCCAGCTGC 480  
QY 481 GCGGCGGCTTCCAGTCCCGAGTGAATATCCGCTGAGGCTGAGGCTGCTGAGGCTGAG 540  
DB 481 GCGGCGGCTTCCAGTCCCGAGTGAATATCCGCTGAGGCTGAGGCTGCTGAGGCTGAG 540  
QY 541 CTGCGCCCTCTGAACTCTGAGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
DB 541 CTGCGCCCTCTGAACTCTGAGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
QY 601 AACATGAGGACCACTGTGCACTGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
DB 601 AACATGAGGACCACTGTGCACTGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
QY 661 GGGGAGGATACCGGCTCTGAGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 720  
DB 661 GGGGAGGATACCGGCTCTGAGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 720  
QY 721 TCGAGGACCACTGTGAAAGGACACCGTTCCCTGCGGAGATCCAGGTGTTCACTGAGC 780  
DB 721 TCGAGGACCACTGTGAAAGGACACCGTTCCCTGCGGAGATCCAGGTGTTCACTGAGC 780  
QY 781 ACCGCTTTGCGAAGTTGACAGAGGCTTGGGAGGCTGCGGAGGCTGAGGCTGAGTGG 840  
DB 781 ACCGCTTTGCGAAGTTGACAGAGGCTTGGGAGGCTGCGGAGGCTGAGGCTGAGTGG 840  
QY 841 GCTTTCTGAGAGGAGGCTCCGAGAAAGAAACAGTGCCTATGAGAGTGTGCTGCTGCTG 900  
DB 841 GCTTTCTGAGAGGAGGCTCCGAGAAAGAAACAGTGCCTATGAGAGTGTGCTGCTGCTG 900  
QY 901 GAAGAATCGCTGAGAGGCTCAGAGACTCAGTCCAGAGCTGAGCAATATCTGACTC 960  
DB 901 GAAGAATCGCTGAGAGGCTCAGAGACTCAGTCCAGAGCTGAGCAATATCTGACTC 960  
QY 961 CTGCTCTGACTTCAAGCCGCTACTTCCAAATATGAGGAGTCTGACTCAACGCGCTGT 1020  
DB 961 CTGCTCTGACTTCAAGCCGCTACTTCCAAATATGAGGAGTCTGACTCAACGCGCTGT 1020  
QY 1021 GCCCAGAGTGTCACTGAGCTGTGTTAACAGACAGTATGCTGAGTGTAAAGCACTC 1080  
DB 1021 GCCCAGAGTGTCACTGAGCTGTGTTAACAGACAGTATGCTGAGTGTAAAGCACTC 1080  
QY 1081 CACACCTCTCTGACACCCCTGTGAGGAGCTGAGTACTCTGAGCTACAGCTGAATTCGA 1140  
DB 1081 CACACCTCTCTGACACCCCTGTGAGGAGCTGAGTACTCTGAGCTACAGCTGAATTCGA 1140

OY	1144	GCGACGCACTTTGAATGGCGAGTGAATTGAGCCCTCCTTCCTCTGTGAGTGACAACG	1200
Dd	1141	GCGACGCACCTTTGAATGGCGAGTGAATTGAGCCCTCCTTCCTCTGTGAGTGACAACG	1200
OY	1201	AGTCTCGGAGCTGTAGAGCAGTCCAGCTGTAATTCCTGCTGTGCTCTGTGACATCCTA	1260
Dd	1201	AGTCTCGGAGCTGTAGAGCAGTCCAGCTGTAATTCCTGCTGTGCTCTGTGACATCCTA	1260
OY	1261	GCCCTGGTTTTTGGGCTCCTTTTTCGTCTCACAGAGGTGGCGTTCTTGTGCAATGAGA	1320
Dd	1261	GCCCTGGTTTTTGGGCTCCTTTTTCGTCTCACAGAGGTGGCGTTCTTGTGCAATGAGA	1320
OY	1321	AGCAGCACAGAAAGGGGAAACCAAAGGGGTTGTAGCTAACGCCACAGAGGTAGCCAG	1380
Dd	1321	AGCAGCACAGAAAGGGGAAACCAAAGGGGTTGTAGCTAACGCCACAGAGGTAGCCAG	1380
OY	1381	ACTGAGAGCTTAGAGGCTGATCTTTGAGAAATGTGAGAAAGCAGCAGAGCATCTGAGGG	1440
Dd	1381	ACTGAGAGCTTAGAGGCTGATCTTTGAGAAATGTGAGAAAGCAGCAGAGCATCTGAGGG	1440
OY	1441	GAGACCGGTAACTGCTCTGCTGCTGCTATATGCACTTCCTTTTACTGCGCAATAAT	1500
Dd	1441	GAGACCGGTAACTGCTCTGCTGCTGCTATATGCACTTCCTTTTACTGCGCAATAAT	1500
OY	1501	TTTTAAATTAATATTATTAAT	1522
Dd	1501	TTTTAAATTAATATTATTAAT	1522
RESULT 7			
AR171392		1522 bp	DNA linear PAT 17-DEC-2001
LOCUS	AR171392		
DEFINITION	Sequence 1 from patent US 6297041.		
ACCESSION	AR171392		
VERSION	AR171392.1 GI:17910342		
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1522)		
TITLE	Zavada, J., Pastorekova, S. and Pastorek, J.		
JOURNAL	MN Gene and protein		
FEATURES	Patent: US 6297041-A 1 02-OCT-2001;		
	Location/Qualifiers		
	1..1522		
	/organism="unknown"		
ORIGIN	/mol_type="unassigned DNA"		
Query Match	100.0%; Score 1522; DB 6; Length 1522;		
Best Local Similarity	100.0%; Pred. No. 3 4e-311;		
Matches 1522; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1	ACAGTACAGCCGATAGCTGCTCCCTGTGCCCCAGCCCTTGAGCTCCCTCTGTGATCCGACC	60
Dd	1	ACAGTACAGCCGATAGCTGCTCCCTGTGCCCCAGCCCTTGAGCTCCCTCTGTGATCCGACC	60
OY	61	CCGTCTCCAGGCTTACTGTGCACTGCTGCTGCACTGCTGCTTCTGATGATCCGACC	120
Dd	61	CCGTCTCCAGGCTTACTGTGCACTGCTGCTGCACTGCTGCTTCTGATGATCCGACC	120
OY	121	CCCAGAGGTTGCCCCGATACAGAGAGATTTCCCTTGGAGAGGCTCTTCTGGGAAA	180
Dd	121	CCCAGAGGTTGCCCCGATACAGAGAGATTTCCCTTGGAGAGGCTCTTCTGGGAAA	180
OY	181	GATGACCCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT	240
Dd	181	GATGACCCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT	240
OY	241	CCACCCGAGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACCT	300
Dd	241	CCACCCGAGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACCT	300

QY	301	GAACTTAAGCTTAATCAGAGAGAGAGGCTCCCTGAAGTTAGAGATCTTACTACTGTT	360
Db	301	GAATTTAAGCTTAATCAGAGAGAGAGGCTCCCTGAAGTTAGAGATCTTACTACTGTT	360
QY	361	GAGGCTCTTGAGATCTTCAGAACCCCAAGATATATGCCACAGGGGCAAAAGAGGGAGT	420
Db	361	GAGGCTCTTGAGATCTTCAGAACCCCAAGATATATGCCACAGGGGCAAAAGAGGGAGT	420
QY	421	GACCAGATGATTGAGCTATGAGAGCGACCCGCTTGCCCCGGGTGTCCCAAGCTGC	480
Db	421	GACCAGATGATTGAGCTATGAGAGCGACCCGCTTGCCCCGGGTGTCCCAAGCTGC	480
QY	481	GCGGCGCGCTTCCAGTCCCCCGGTGGATATCCGCCCCCAGCTCGCGCTTTCTGCGCGGC	540
Db	481	GCGGCGCGCTTCCAGTCCCCCGGTGGATATCCGCCCCCAGCTCGCGCTTTCTGCGCGGC	540
QY	541	CTGCGCCCCCTGGAATCCTTGAGCTTCCAGCTCCCGCGCTCCAGAACCTGCCTGCGC	600
Db	541	CTGCGCCCCCTGGAATCCTTGAGCTTCCAGCTCCCGCGCTCCAGAACCTGCCTGCGC	600
QY	601	AACATATGCCACAGTGTGCATTTGACCTGCTCTTGAGATATGATGCTTGAGTATGCC	660
Db	601	AACATATGCCACAGTGTGCATTTGACCTGCTCTTGAGATATGATGCTTGAGTATGCC	660
QY	661	GGGCGGAGATACCGGCTCTGAGCTGCATCTGCACCTGAGGAGCGTGAAGTCTGCGGAC	720
Db	661	GGGCGGAGATACCGGCTCTGAGCTGCATCTGCACCTGAGGAGCGTGAAGTCTGCGGAC	720
QY	721	TCCGAGCAACTGTGGAAGGCCACCGTTCCCTGCGCAGATATCCAGTGTTCACCTGACG	780
Db	721	TCCGAGCAACTGTGGAAGGCCACCGTTCCCTGCGCAGATATCCAGTGTTCACCTGACG	780
QY	781	ACGCGCTTTGCAAGATTGAAGAGGCGCTTGAGGAGCGCCCGAGAGCGCTGCTTGAGGC	840
Db	781	ACGCGCTTTGCAAGATTGAAGAGGCGCTTGAGGAGCGCCCGAGAGCGCTGCTTGAGGC	840
QY	841	GCTTTCTGAGAGAGGCGCCCGAGAGAAACAGTGCCTTATAGCAGATTGTCTGCTGCTTG	900
Db	841	GCTTTCTGAGAGAGGCGCCCGAGAGAAACAGTGCCTTATAGCAGATTGTCTGCTGCTTG	900
QY	901	GAGGAATATGCTGAGAGAGGCTCAGAGATCAGGTCCCAAGATTCATATCTGACATC	960
Db	901	GAGGAATATGCTGAGAGAGGCTCAGAGATCAGGTCCCAAGATTCATATCTGACATC	960
QY	961	CTAGCCTCTGACTTACGCGCTACTTCCCATATATAGAGGAGTCTGACTACACGCGCTGT	1020
Db	961	CTAGCCTCTGACTTACGCGCTACTTCCCATATATAGAGGAGTCTGACTACACGCGCTGT	1020
QY	1021	GCCAGAGGTATCTGACTGTGTTTACCAGACAGTGAATGTGAGTCTTAAAGACGCTC	1080
Db	1021	GCCAGAGGTATCTGACTGTGTTTACCAGACAGTGAATGTGAGTCTTAAAGACGCTC	1080
QY	1081	CACACCCTCTGACACCCCTGAGGAGACCTGAGTACTCTGAGCTACAGCTGAATTCGCA	1140
Db	1081	CACACCCTCTGACACCCCTGAGGAGACCTGAGTACTCTGAGCTACAGCTGAATTCGCA	1140
QY	1141	GCGACGACCTCTTGAATGAGCGCAGATATGAGGCGCTCTCTCTGATGAGTGAACAGC	1200
Db	1141	GCGACGACCTCTTGAATGAGCGCAGATATGAGGCGCTCTCTCTGATGAGTGAACAGC	1200
QY	1201	AGTCTCTGAGGCTCTGAGCTCAGTCAAGTGAATTCCTGCTGAGCTGTGTATCAATCTTA	1260
Db	1201	AGTCTCTGAGGCTCTGAGCTCAGTCAAGTGAATTCCTGCTGAGCTGTGTATCAATCTTA	1260
QY	1261	GCCCTGTTTTGACCTCTTTTGTCTGTCACACAGAGCTCGCGTCTCTTGACAGATAGGA	1320
Db	1261	GCCCTGTTTTGACCTCTTTTGTCTGTCACACAGAGCTCGCGTCTCTTGACAGATAGGA	1320
QY	1321	AGGACACATAGAGGGGAACTAAAGGGGTGTAGCTACCGCCACAGAGATTAACCCAG	1380
Db	1321	AGGACACATAGAGGGGAACTAAAGGGGTGTAGCTACCGCCACAGAGATTAACCCAG	1380
QY	1381	ACTGAGGCTTAAGGCTGATCTTGAGGAATGTGAAGACCAACCGAGAGCACTTGAGAGG	1440

Db	1381	ACTGAGACCTTAAGGCTGATCTTGGAAATCTGAAAGCCAGCCAGGACATCTGAGG	1440
QY	1441	GGAGCCGGPACTGTCTCTGCTCTCTCATTAAGCACTTCTTTTAACGCCAAGAAAT	1500
Db	1441	GGAGCCGGPAACTGCTCTCTCTCATTAAGCACTTCTTTTAACGCCAAGAAAT	1500
QY	1501	TTTTAAATTAATATTTAAT	1522
Db	1501	TTTTAAATTAATATTTAAT	1522
RESULT 8			
LOCUS	AA171563	1522 bp	linear
DEFINITION	Sequence 1 from patent US 6297051.		PAT 17-DEC-2001
ACCESSION	AA171563		
VERSION	AA171563.1	GI:17910513	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1522)		
TITLE	Zavada, V., Pastorekova, S. and Pastorek, J.		
JOURNAL	MN gene and protein		
FEATURES	Patent: US 6297051-A 1 02-OCT-2001;		
	Location/Qualifiers		
	1..1522		
ORIGIN	/organism="unknown"		
	/mol_type="unassigned DNA"		
Query Match	100.0%; Score 1522; DB 6; Length 1522;		
Best Local Similarity	100.0%; Pred. No. 3,4e-11;		
Matches 1522; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	ACAGTAGCCGATGCTCCCTGTGCCCCAGCCCTTGGCTCCCTCTGTATCCCGGC	60
Db	1	ACAGTAGCCGATGCTCCCTGTGCCCCAGCCCTTGGCTCCCTCTGTATCCCGGC	60
QY	61	CTGTCTCAGGCTCACTGTGCATCTGTCTGTCTGTCTGTATGCTGTGAT	120
Db	61	CTGTCTCAGGCTCACTGTGCATCTGTCTGTCTGTCTGTATGCTGTGAT	120
QY	121	CCCCAAGGTGCCCCGATGAGAGGATATCCCCCTTGGAGAGAGGCTCTTGAGAA	180
Db	121	CCCCAAGGTGCCCCGATGAGAGGATATCCCCCTTGGAGAGAGGCTCTTGAGAA	180
QY	181	GATGACCACTGGGCGAGAGGATCTGCCAGTGAAGAGATTCAACCAAGAGAGAT	240
Db	181	GATGACCACTGGGCGAGAGGATCTGCCAGTGAAGAGATTCAACCAAGAGAGAT	240
QY	241	CCACCCGAGAGAGGATTTACCTGGAAGAGATTTACCTGGAAGAGATTTACTT	300
Db	241	CCACCCGAGAGAGGATTTACCTGGAAGAGATTTACCTGGAAGAGATTTACTT	300
QY	301	GAAATTACCTTAATCAGAGAGAGGCTCCTCTGAAGTTAGAGATTTACTTCTGT	360
Db	301	GAAATTACCTTAATCAGAGAGAGGCTCCTCTGAAGTTAGAGATTTACTTCTGT	360
QY	361	GAGGCTCTGAGATCTCAAGAACCCAGAAATATCCACAGAGGACAAAGAGGAT	420
Db	361	GAGGCTCTGAGATCTCAAGAACCCAGAAATATCCACAGAGGACAAAGAGGAT	420
QY	421	GACCAAGTATTTGGGCTATGAGAGGCACTCGCTTGACCCCGGATGTCCTCAGCTGC	480
Db	421	GACCAAGTATTTGGGCTATGAGAGGCACTCGCTTGACCCCGGATGTCCTCAGCTGC	480
QY	481	GGGGGCGCTTCAAGTCCCGGTGATATCGGCCCAAGTCTGCGCGCTTCTGCGCGGC	540
Db	481	GGGGGCGCTTCAAGTCCCGGTGATATCGGCCCAAGTCTGCGCGCTTCTGCGCGGC	540
QY	541	CTGCGCCCTGAACTCTTGAGCTTCAAGTCCCGGCTCCCAAACTGGGCTGGC	600

Db	541	CTGCGCCCCCTGGAACTCCTGGGCTTCCAGCTCCCCCGCTCCCAAGACTGGGCTGGCC	600
QY	601	AACAAATGGCCACAGTGTSCAACTGACCTGCTCTGGGCTAGAGATGSCCTTGGGTGCC	660
Db	601	AAACAATGGCCACAGTGTSCAACTGACCTGCTCTGGGCTAGAGATGSCCTTGGGTGCC	660
QY	661	GAGCGGGAATACCGGAGCTCTGAGGTGATCTGCACTGGGGGCTTCAGATGCTCGGAGC	720
Db	661	GAGCGGGAATACCGGAGCTCTGAGGTGATCTGCACTGGGGGCTTCAGATGCTCGGAGC	720
QY	721	TGCGAGCACTGTGTGAAGGCCAACCGTTTCCCTGCGAGATCCAGTGGTTCACTCAAC	780
Db	721	TGCGAGCACTGTGTGAAGGCCAACCGTTTCCCTGCGAGATCCAGTGGTTCACTCAAC	780
QY	781	ACGCGCTTTGGCAGAGTTGACGAGGCTTGGGGGCGCCGAGAGGCTGGCCGCTGTGTGAC	840
Db	781	ACGCGCTTTGGCAGAGTTGACGAGGCTTGGGGGCGCCGAGAGGCTGGCCGCTGTGTGAC	840
QY	841	GCTTTCTTGAGAGAGGGCGCCGGAAGAAACAGTGCCTATGAGCATTTGCTGTGGCTTG	900
Db	841	GCTTTCTTGAGAGAGGGCGCCGGAAGAAACAGTGCCTATGAGCATTTGCTGTGGCTTG	900
QY	901	GAAGAAATGCTGAGAGAAAGGCTCAGAGCTCAGGTCACAGACTGGACATATCTGACTC	960
Db	901	GAAGAAATGCTGAGAGAAAGGCTCAGAGCTCAGGTCACAGACTGGACATATCTGACTC	960
QY	961	CTGCGCTCTGACTTTCAGCGGCTCATTCCTCAATGAGGGGCTGTGACTCAACCGGCTGT	1020
Db	961	CTGCGCTCTGACTTTCAGCGGCTCATTCCTCAATGAGGGGCTGTGACTCAACCGGCTGT	1020
QY	1021	GCCCAAGGTGTATCTGAGCTGTGTATTAACCAACAGTATGCTGAGTCTTAAGCACTC	1080
Db	1021	GCCCAAGGTGTATCTGAGCTGTGTATTAACCAACAGTATGCTGAGTCTTAAGCACTC	1080
QY	1081	CACACCTCTCTGACACCTCTGGGGACCTGGTGAACCTCGAGTCAAGTGAACCTCCGA	1140
Db	1081	CACACCTCTCTGACACCTCTGGGGACCTGGTGAACCTCGAGTCAAGTGAACCTCCGA	1140
QY	1141	GCGACGACCCCTTTGAATGGGCGAGTATGAGGCTCTTCCCTGCTGGAATGACAGC	1200
Db	1141	GCGACGACCCCTTTGAATGGGCGAGTATGAGGCTCTTCCCTGCTGGAATGACAGC	1200
QY	1201	AGTCTCTGGGCTCTGAGACCAAGTCAAGTGAATTCCTGCGCTGACTCTGTATCAATCCGA	1260
Db	1201	AGTCTCTGGGCTCTGAGACCAAGTCAAGTGAATTCCTGCGCTGACTCTGTATCAATCCGA	1260
QY	1261	GCCCTGTATTTGGGCTCTCTTTTGTCTGTACAAGGCTGGCGTTCTTGTGCAATGAGA	1320
Db	1261	GCCCTGTATTTGGGCTCTCTTTTGTCTGTACAAGGCTGGCGTTCTTGTGCAATGAGA	1320
QY	1321	AGCAGACACAGAAAGGGGAACAAAGGGGGTGTGAGCTACCGCCACAGACAGATTAACGAG	1380
Db	1321	AGCAGACACAGAAAGGGGAACAAAGGGGGTGTGAGCTACCGCCACAGACAGATTAACGAG	1380
QY	1381	ACTGAGCCTTAAGGCTTGATCTTGGAGATGTGAGAGGCCAGCCAGAGGCACTTGAAGG	1440
Db	1381	ACTGAGCCTTAAGGCTTGATCTTGGAGATGTGAGAGGCCAGCCAGAGGCACTTGAAGG	1440
QY	1441	GAGCGCGTAACTGTGCTCTGTCTGATTAATGCACTTCTTTAACTGCAGAAATAAT	1500
Db	1441	GAGCGCGTAACTGTGCTCTGTCTGATTAATGCACTTCTTTAACTGCAGAAATAAT	1500
QY	1501	TTTTAAATAATTAATTTTAAAT	1522
Db	1501	TTTTAAATAATTAATTTTAAAT	1522







```

FEATURES
  Source
    Location/Qualifiers
      1..1552
      /organism="Homo sapiens"
      /mol_type="unassigned DNA"
      /db_xref="taxon:9606"
  ORIGIN
    Query Match
    Best Local Similarity 100.0%; Score 1522; DB 6; Length 1552;
    Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  Db 931 GAAGAAATGCTGAGGAAGGCTCAGAGACTCAGGCTCCAGAGACTGAGCATATCTGCACTC 990
  Qy 961 CTGCGCTGACTGACCGGCTACTTCCAAATATAGAGGGGTCTGACTACACCGCCCTGT 1020
  Db 991 CTGCGCTGACTGACCGGCTACTTCCAAATATAGAGGGGTCTGACTACACCGCCCTGT 1050
  Qy 1021 GCCCAGGGTGTATCTGACCTGTGTTTAAACGACAGATGATGCTGAGTCTTAACAGCTC 1080
  Db 1051 GCCCAGGGTGTATCTGACCTGTGTTTAAACGACAGATGATGCTGAGTCTTAACAGCTC 1110
  Qy 1081 CACACCTCTGACACCTGTGAGGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
  Db 1111 CACACCTCTGACACCTGTGAGGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1170
  Qy 1141 GCGACGACCTTTGATAGGAGATGATTTGAGGCTCTCTCTGCTGAGTGAACAGC 1200
  Db 1171 GCGACGACCTTTGATAGGAGATGATTTGAGGCTCTCTCTGCTGAGTGAACAGC 1230
  Qy 1201 AGTCTCGGGCTGCTGAGCAGTCCAGCTGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
  Db 1231 AGTCTCGGGCTGCTGAGCAGTCCAGCTGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1290
  Qy 1261 GCGCTGCTTTTGGCTCTCTTTTGTCTGCTCAGAGCTGCGCTTCTTGTGCAATGAGA 1320
  Db 1291 GCGCTGCTTTTGGCTCTCTTTTGTCTGCTCAGAGCTGCGCTTCTTGTGCAATGAGA 1350
  Qy 1321 AGGACGACAGAGGAGGAAACCAAGGGGTGTGAGCTACCGCCAGCAGAGGATGAGCGAG 1380
  Db 1351 AGGACGACAGAGGAGGAAACCAAGGGGTGTGAGCTACCGCCAGCAGAGGATGAGCGAG 1410
  Qy 1381 ACTGAGCTTAGAGGCTGAGTCTTGTGAGATGTGAGAAAGCAAGCAAGAGGATTTGAGGG 1440
  Db 1411 ACTGAGCTTAGAGGCTGAGTCTTGTGAGATGTGAGAAAGCAAGCAAGAGGATTTGAGGG 1470
  Qy 1441 GGAGCGGGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
  Db 1471 GGAGCGGGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1530
  Qy 1501 TTTTAAATTAATTAATTAAT 1522
  Db 1531 TTTTAAATTAATTAATTAAT 1552

  RESULT 12
  AX333244 1552 bp DNA linear PAT 09-JAN-2002
  LOCUS AX333244
  DEFINITION Sequence 3753 from Patent WO0194629.
  ACCESSION AX333244
  VERSION AX333244.1 GI:18123878
  KEYWORDS
  SOURCE Homo sapiens (human)
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

  REFERENCE
  1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
    Horrigan, S., Soppet, D.R. and Weaver, Z.
    Cancer gene determination and therapeutic screening using signature
    gene sets
    Patent: WO 0194629-A 3753 13-DEC-2001;
    Avalon Pharmaceuticals (US)

  JOURNAL
    Location/Qualifiers
      1..1552
      /organism="Homo sapiens"
      /mol_type="unassigned DNA"
      /db_xref="taxon:9606"
  ORIGIN
    Query Match
    Best Local Similarity 100.0%; Score 1522; DB 6; Length 1552;
    Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  Db 901 GAAGAAATGCTGAGGAAGGCTCAGAGACTCAGGCTCCAGAGACTGAGCATATCTGCACTC 960

```





Db	211	GATGACCACTGGGCGAGAGGATCTGGCCAGTGAAGATTCACCGAGAGGAGAT	270
Qy	241	CCACCCGGAGAGAGATTTACTTGGAGAGAGATCTACCTGAGAGAGATCTACT	300
Db	271	CCACCCGGAGAGAGATTTACTTGGAGAGAGATTTACTTGGAGAGAGATCTACT	330
Qy	301	GAAGTTAACCTTAAATCAGAGAGAGAGGCTCCCTGAAGTTAAGAGATCTACTGTT	360
Db	331	GAAGTTAAGCTTAAATCAGAGAGAGAGGCTCCCTGAAGTTAAGAGATCTACTGTT	390
Qy	361	GAGGCTCTTGGAGATCTTCAAGAACCCCAAAATATGCTCCACAGGGACAAAGAGGAT	420
Db	391	GAGGCTCTTGGAGATCTTCAAGAACCCCAAGATATATGCTCCACAGGACAAAGAGGAT	450
Qy	421	GACCAAGTCAATTGGCGCTATGAGAGCGACCCGCGCTGGCCCGGAGTGTCCAGCAGTC	480
Db	451	GACCAAGTCAATTGGCGCTATGAGAGCGACCCGCGCTGGCCCGGAGTGTCCAGCAGTC	510
Qy	481	GCGGGCGGCTTCCAGTCCCGGCTGGATATCCGCCCCAGGTGCGCGCTTTTGGCCCGGC	540
Db	511	GCGGGCGGCTTCCAGTCCCGGCTGGATATCCGCCCCAGGTGCGCGCTTTTGGCCCGGC	570
Qy	541	CTGCGCCCCCTGAACTCCTGGCTTTCAGCTCCCGCGCTCCAGAACCTGGCGTGGAC	600
Db	571	CTGCGCCCCCTGAACTCCTGGCTTTCAGCTCCCGCGCTCCAGAACCTGGCGTGGAC	630
Qy	601	AACAAAGGCAACAGTGTGCAACTGACCCCTGCTCTGGGTAAAGATGTGCTGGGTCCC	660
Db	631	AACAAAGGCAACAGTGTGCAACTGACCCCTGCTCTGGGTAAAGATGTGCTGGGTCCC	690
Qy	661	GGGCGGGAGTACCGGACTCTGAGCTGACTGCACTGGGGGGCTGACAGTGGTCCGGAC	720
Db	691	GGGCGGGAGTACCGGACTCTGAGCTGACTGCACTGGGGGGCTGACAGTGGTCCGGAC	750
Qy	721	TGGAGCAACACTGTGGAAGGCCACCGTTCCCTGCGCGAATCCAGCTGGTTCACTCAGC	780
Db	751	TGGAGCAACACTGTGGAAGGCCACCGTTCCCTGCGCGAATCCAGCTGGTTCACTCAGC	810
Qy	781	ACGCGCTTTGCCAGAGTTGACGAGGCTTTGGGGGCGCCCGGAGGCTTGGCCGTGTGTGGCC	840
Db	811	ACGCGCTTTGCCAGAGTTGACGAGGCTTTGGGGGCGCCCGGAGGCTTGGCCGTGTGTGGCC	870
Qy	841	GCGTTTCTGGAAGAGGGGCCGGAAGAAAACAGTSCCTATAGAGAGTGTGCTGTGCTTGG	900
Db	871	GCGTTTCTGGAAGAGGGGCCGGAAGAAAACAGTSCCTATAGAGAGTGTGCTGTGCTTGG	930
Qy	901	GAAGAAATCGCTAGAGAAAGCTCAGACCTCAGTCCAGACTGACATATCTGACTC	960
Db	931	GAAGAAATCGCTAGAGAAAGCTCAGACCTCAGTCCAGACTGACATATCTGACTC	990
Qy	961	CTGCGCTCTGACTCAGCGCGCTACTTCCCAATATAGAGGGGTCTGTGACTCACCGGCCGT	1020
Db	991	CTGCGCTCTGACTCAGCGCGCTACTTCCCAATATAGAGGGGTCTGTGACTCACCGGCCGT	1050
Qy	1021	GCCCAAGGTGTCACTGTGACTGTGTTAACCAACAGCACTGATGTGAGTGTCTAAGCAGCTC	1080
Db	1051	GCCCAAGGTGTCACTGTGACTGTGTTAACCAACAGCACTGATGTGAGTGTCTAAGCAGCTC	1110
Qy	1081	CACACCTCTGTGACACCCCTGTGGGGACCTGTGTGACTCTGTGGCTTACAGTGAACCTCCG	1140
Db	1111	CACACCTCTGTGACACCCCTGTGGGGACCTGTGTGACTCTGTGGCTTACAGTGAACCTCCG	1170
Qy	1141	GCGAGCGAGCCTTTGATATGGCGAGTGAATTGAGGCTCTCTTCCCTGTGAGTGAACAGC	1200
Db	1171	GCGAGCGAGCCTTTGATATGGCGAGTGAATTGAGGCTCTCTTCCCTGTGAGTGAACAGC	1230
Qy	1201	AGTCTCTGGGCTGTGAGACAGTCCAGTGAATTCGTGCGGTGCTGTGTGACATCTCA	1260
Db	1231	AGTCTCTGGGCTGTGAGACAGTCCAGTGAATTCGTGCGGTGCTGTGTGACATCTCA	1290
Qy	1261	GCCCTGTGTTTGGCTCTCTTTTGTGTGTGACAGGCTGCGCTTGTGTGACATGAGA	1320

Dp		1291	GCCCTGATTTTGGCCTCTCTTTTGCTGTACACAGGCGTGGCTTCCTTGAGAGTAGAGA	1380
Oy		1321	AGGCACACAGAAAGGGGAACCAAAAGGGGGGTGTAGCTACTACCCTCACAGAGGTACCCAG	1380
Dp		1351	AGGCACACAGAAAGGGGAACCAAAAGGGGGGTGTAGCTACTACCCTCACAGAGGTACCCAG	1410
Oy		1381	ACTGAGACCTTAAGAGCTGTGATCTTGGAGAAATGTGAGAACCCAGCAGAGCATCTGAGGG	1440
Dp		1411	ACTGAGACCTTAAGAGCTGTGATCTTGGAGAAATGTGAGAACCCAGCAGAGCATCTGAGGG	1470
Oy		1441	GAGGCCGGTAACTGTCTCTGTCTCTCTCTATTATGCGCATCTCTTTTAACTGCCAAGAAATT	1500
Dp		1471	GAGGCCGGTAACTGTCTCTGTCTCTCTCTATTATGCGCATCTCTTTTAACTGCCAAGAAATT	1530
Oy		1501	TTTTAAATAATTAATTTATTAAT	1522
Dp		1531	TTTTAAATAATTAATTTATTAAT	1552
RESULT 14				
LOCUS	HSMATUMN	1552 bp	mRNA	linear PRI 10-OCT-1995
DEFINITION	H.sapiens Matru MN mRNA for p54/58n protein.			
ACCESSION	X66839			
VERSION	X66839.1 GI:1000701			
KEYWORDS	transmembrane glycoprotein.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
TITLE	1 (bases 1 to 1552)			
JOURNAL	Pastorek,J. Direct Submission Submitted (11-JUN-1992) J. Pastorek, Institute of Virology, Slovak Academy of Sciences, Dubravskva 9, 842 46 Bratislava, SLOVAK REPUBLIC			
REMARK	revised by [3] MAT			
REFERENCE	2 (bases 1 to 1552)			
AUTHORS	Pastorek,J., Pastorekova,S., Callebaut,I., Mornon,J., Zelnik,V., Opavsky,R., Zatovicova,M., Liao,S., Portetelle,D., Standeride,E.J., Zavada,J. and Burny,A. Cloning and characterization of MN, a human tumor-associated protein with a domain homologous to carbonic anhydrase and a putative helix-loop-helix DNA binding segment			
TITLE	Oncogene 9 (10), 2877-2888 (1994)			
JOURNAL	Medline			
REFERENCE	PIMED			
AUTHORS	3 (bases 1 to 1552)			
TITLE	Pastorek,J. Direct Submission Submitted (19-JUL-1994) J. Pastorek, Institute of Virology, Slovak Academy of Sciences, Dubravskva 9, 842 46 Bratislava, SLOVAK REPUBLIC			
REMARK	revised by [4] MAT			
REFERENCE	4 (bases 1 to 1552)			
AUTHORS	Pastorek,J. Direct Submission Submitted (28-SEP-1995) J. Pastorek, Institute of Virology, Slovak Academy of Sciences, Dubravskva 9, 842 46 Bratislava, SLOVAK REPUBLIC			
COMMENT	On Sep 29, 1995 this sequence version replaced gi:558593.			
FEATURES	Location/Qualifiers			
source	1..1552			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	1..1519			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="MN1"			
	/cell_type="epithelial"			
	/tissue_type="carcinoma"			





```
QY      841 GCCTTCTGAGAGAGGCGCCGGAGAAAACAGTGCCTATGAGCAGTTGCTGCTGCTTG 900
      |||
Db      933 GCCTTCTGAGAGAGGCGCCGGAGAAAACAGTGCCTATGAGCAGTTGCTGCTGCTTG 992
      |||
QY      901 GAGAAATCGCTGAGGAAGCTCAGAGACTCAGAGTCCAGAGCTGACATATCTGCATC 960
      |||
Db      993 GAGAAATCGCTGAGGAAGCTCAGAGACTCAGAGTCCAGAGCTGACATATCTGCATC 1052
      |||
QY      961 CTGCCCTCTGACTGAGCGCTACTTCCAAATATAGAGGGTCTGACTACACCGCCCTGT 1020
      |||
Db      1053 CTGCCCTCTGACTGAGCGCTACTTCCAAATATAGAGGGTCTGACTACACCGCCCTGT 1112
      |||
QY      1021 GCCCAGGTCATCTGAGTCTGTGTTTAAACCAGACAGTGAATGCTGAGTCTTAAGCAGCTC 1080
      |||
Db      1113 GCCCAGGTCATCTGAGTCTGTGTTTAAACCAGACAGTGAATGCTGAGTCTTAAGCAGCTC 1172
      |||
QY      1081 CACACCCCTCTGAGACACCTGTGGGGAACCTGTGACTCTCGGCTACAGCTGAACCTCCGA 1140
      |||
Db      1173 CACACCCCTCTGAGACACCTGTGGGGAACCTGTGACTCTCGGCTACAGCTGAACCTCCGA 1232
      |||
QY      1141 GCGACGACGCTTTGAATGGCGAGTGAATTGAGGCTCTTCCCTGCTGAGTGAACAGC 1200
      |||
Db      1233 GCGACGACGCTTTGAATGGCGAGTGAATTGAGGCTCTTCCCTGCTGAGTGAACAGC 1292
      |||
QY      1201 AGTCTCGGGCTGCTGAGCCAGTCCAGCTGAATTCCTGCTGCTGCTGCTGCTGCTGCTG 1260
      |||
Db      1293 AGTCTCGGGCTGCTGAGCCAGTCCAGCTGAATTCCTGCTGCTGCTGCTGCTGCTGCTG 1352
      |||
QY      1261 GCCCTGATTTTGGGCTCCTTTTGTCTGTCAACAGCGTGGGTTCTTGTGCAATGAGA 1320
      |||
Db      1353 GCCCTGATTTTGGGCTCCTTTTGTCTGTCAACAGCGTGGGTTCTTGTGCAATGAGA 1412
      |||
QY      1321 AGGCAGCACAGAAAGGGGAAACCAAGGGGGTGTGAGCTACGCGCCAGCAGAGTGAACCGAG 1380
      |||
Db      1413 AGGCAGCACAGAAAGGGGAAACCAAGGGGGTGTGAGCTACGCGCCAGCAGAGTGAACCGAG 1472
      |||
QY      1381 ACTGAGCCTAGAGGCTGATCTTGAGAAATGTGAGAAAGCCAGCCAGAGGATCTGAGGG 1440
      |||
Db      1473 ACTGAGCCTAGAGGCTGATCTTGAGAAATGTGAGAAAGCCAGCCAGAGGATCTGAGGG 1532
      |||
QY      1441 GAGGCGGTAAGTGTCTGTCTGCTCAATTAGCCACTTCTTTTAACTGCCAAGAAATT 1500
      |||
Db      1533 GAGGCGGTAAGTGTCTGTCTGCTCAATTAGCCACTTCTTTTAACTGCCAAGAAATT 1592
      |||
QY      1501 TTTTAAATTAATTTTAAAT 1522
      |||
Db      1593 TTTTAAATTAATTTTAAAT 1614
      |||
```

Search completed: March 7, 2004, 04:36:17  
Job time : 4053 secs



PT in the detection of MN antigens and antibodies, and in the treatment of  
 PT (pre)neoplastic disease.

PS Claim 1; Fig 1; 102pp; English.

XX The present sequence is the full length Mutu endogenous cellular  
 CC component, MN, cDNA clone, which was isolated from lymphocytic  
 CC choriomeningitis virus (LCMV) infected Hela cells. Persistent LCMV, the  
 CC exogenous Mutu transmissible agent (MX), infection increases the  
 CC expression level of the MN gene. MN is a putative oncogene, and can  
 CC therefore be used in the development of probe, for the diagnosis and  
 CC treatment of neoplastic (NP), or pre-NP diseases. NP diseases can be  
 CC treated using DNA antisense to MN transcribed mRNA, anti-MN protein  
 CC antibodies can be used for the diagnosis NP or pre-NP diseases and a  
 CC vaccine contg. immunogenic amounts of the MN protein can be used to  
 CC immunise a vertebrate against a NP disease associated with MN antigen  
 CC expression. (updated on 25-MAR-2003 to correct PR field.)  
 CC  
 XX

Sequence 1522 BP; 297 A; 458 C; 453 G; 314 T; 0 U; 0 Other;

Query Match 100.0%; Score 1522; DB 2; Length 1522;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ACAGTCAGCCGATGCTCCCTGTGCGCCAGCCCTGAGCTCCCTGTGTATCCCGGCC	60
DB	1	ACAGTCAGCCGATGCTCCCTGTGCGCCAGCCCTGAGCTCCCTGTGTATCCCGGCC	60
QY	61	CTGTCTCAAGGCTCACTGTGCAATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT	120
DB	61	CTGTCTCAAGGCTCACTGTGCAATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT	120
QY	121	CCCCAGAGTTGCCCCGGATGAGAGATTTCCCTTGGAGAGAGCTCTTGGGAA	180
DB	121	CCCCAGAGTTGCCCCGGATGAGAGATTTCCCTTGGAGAGAGCTCTTGGGAA	180
QY	181	GATGACCCACTGAGGCTGAGAGATCTGCCAGTGAAGAGATTTCAACCAAGAGAGAT	240
DB	181	GATGACCCACTGAGGCTGAGAGATCTGCCAGTGAAGAGATTTCAACCAAGAGAGAT	240
QY	241	CCACCCGAGAGAGATCTGAGAGAGATTTCACTGAGAGAGAGATTTACT	300
DB	241	CCACCCGAGAGAGATCTGAGAGAGATTTCACTGAGAGAGAGATTTACT	300
QY	301	GAAATTAGCTTAATCAGAAAGAGGCTCCCTGAGTGAAGATCTACTTCT	360
DB	301	GAAATTAGCTTAATCAGAAAGAGGCTCCCTGAGTGAAGATCTACTTCT	360
QY	361	GAGGCTCTGAGATCTCTCAAGAAACCCAGAAATATGCCAGAGAGAGAGGAT	420
DB	361	GAGGCTCTGAGATCTCTCAAGAAACCCAGAAATATGCCAGAGAGAGAGGAT	420
QY	421	GACCAAGATCTGAGGCTATGAGAGGCAACCCGCTCTGAGCCCGGATGTCCCAAGCTGC	480
DB	421	GACCAAGATCTGAGGCTATGAGAGGCAACCCGCTCTGAGCCCGGATGTCCCAAGCTGC	480
QY	481	GCGGCGCGCTTCAAGTCCCGGTGAGATTCGCGCCCAAGTCTGCGCCCGGCC	540
DB	481	GCGGCGCGCTTCAAGTCCCGGTGAGATTCGCGCCCAAGTCTGCGCCCGGCC	540
QY	541	CTGCGCCCTTGAAGTCTGAGGCTTCAAGCTCCCGCTCCCAAGTCTGCGCCCGGCC	600
DB	541	CTGCGCCCTTGAAGTCTGAGGCTTCAAGCTCCCGCTCCCAAGTCTGCGCCCGGCC	600
QY	601	AAAGATGAGCAAGTGTGCAACCTGAGTCTGAGGCTGAGATGAGTCTGTGATGCC	660
DB	601	AAAGATGAGCAAGTGTGCAACCTGAGTCTGAGGCTGAGATGAGTCTGTGATGCC	660
QY	661	GAGCGGAGATACCGGAGCTGTGAGAGCTGCACTGTGAGGAGGAGTGTGCGGAG	720
DB	661	GAGCGGAGATACCGGAGCTGTGAGAGCTGCACTGTGAGGAGGAGTGTGCGGAG	720
QY	721	TGGAAGCACTGTGAGAGGCAACGTTTCCCTGCGGAGATCCAGTGTGATCACTCAAG	780

DB	721	TCGAGACACACTGTGAGAGGCGCACCGTTTCCCTCCAGATTCACAGTGTTCACCTCAGC	780
QY	781	ACCGCTTTGACAGAGTGTGAGAGGCTTGGGAGCGCCGAGAGGCTGTGCGATTTGGCC	840
DB	781	ACCGCTTTGACAGAGTGTGAGAGGCTTGGGAGCGCCGAGAGGCTGTGCGATTTGGCC	840
QY	841	GCTTTCTGAGAGAGGCGCCGAGAGAAACAGTCTCTATGAGCATGTTGTCTGCTTG	900
DB	841	GCTTTCTGAGAGAGGCGCCGAGAGAAACAGTCTCTATGAGCATGTTGTCTGCTTG	900
QY	901	GAGAAATGCTGAGAGAGGCTCAAGACTAGGCTCCAGGATCCAGATCATATGCACTC	960
DB	901	GAGAAATGCTGAGAGAGGCTCAAGACTAGGCTCCAGGATCCAGATCATATGCACTC	960
QY	961	CTGCGCTTGAATCTGAGGCGCTACTTCCATATGAGAGGCTGTGACTCAACCGCTGT	1020
DB	961	CTGCGCTTGAATCTGAGGCGCTACTTCCATATGAGAGGCTGTGACTCAACCGCTGT	1020
QY	1021	GCCAGAGGATGATGAGACTGTGTTTACAGACAGTGAATGAGTGTGAGTGTGAGAGCTC	1080
DB	1021	GCCAGAGGATGATGAGACTGTGTTTACAGACAGTGAATGAGTGTGAGTGTGAGAGCTC	1080
QY	1081	CACACCTCTGACACCTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG	1140
DB	1081	CACACCTCTGACACCTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG	1140
QY	1141	GCGAGGAGCTTTGAAATGGGAGATTTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCT	1200
DB	1141	GCGAGGAGCTTTGAAATGGGAGATTTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCT	1200
QY	1201	AGTCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTG	1260
DB	1201	AGTCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTG	1260
QY	1261	GCTCTGTGTTTGGGCTCTCTTTTGTGCTGACAGGCTGTGCTCTTGTGCAATGAGA	1320
DB	1261	GCTCTGTGTTTGGGCTCTCTTTTGTGCTGACAGGCTGTGCTCTTGTGCAATGAGA	1320
QY	1321	AGGAGCAGAGAGGAGGAGACCAAGAGGAGTGTGAGTACCGCCAGAGAGAGTACCG	1380
DB	1321	AGGAGCAGAGAGGAGGAGACCAAGAGGAGTGTGAGTACCGCCAGAGAGAGTACCG	1380
QY	1381	ACTGAGCTTGAAGGCTGTGATTTGAGAAATGTGAGAACCCAGAGGATCTGAGGG	1440
DB	1381	ACTGAGCTTGAAGGCTGTGATTTGAGAAATGTGAGAACCCAGAGGATCTGAGGG	1440
QY	1441	GAGCGGATGATGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT	1500
DB	1441	GAGCGGATGATGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT	1500
QY	1501	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1522
DB	1501	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1522

# RESULT 2

AA16540 ID AA16540 standard; cDNA; 1522 BP.

AA16540; 16-JUN-2000 (first entry)

DE Human MN protein encoding cDNA seq ID NO:1.

XX Human; MN protein; MN gene; oncogene; carbonic anhydrase; tumour;

XX oncogenesis; diagnosis; neoplastic disease; cancer; carcinoma;

XX MN/CA IX isoenzyme; ds.

OS Homo sapiens.

XX US6027887-A.



XX 22-FEB-2000.  
XX 24-JAN-1997; 97US-00787739.  
XX 21-OCT-1992; 92US-00964589.  
XX 30-DEC-1993; 93US-00177093.  
XX 15-JUN-1994; 94US-00260190.  
XX 07-JUN-1995; 95US-00477504.  
XX 07-JUN-1995; 95US-00481658.  
XX 07-JUN-1995; 95US-00485049.  
XX 07-JUN-1995; 95US-00485862.  
XX 07-JUN-1995; 95US-00485863.  
XX 07-JUN-1995; 95US-00486755.  
XX 07-JUN-1995; 95US-00487077.  
XX (SLSC-) SLOVAK ACAD SCI INST VIROLOGY.  
XX Pastorek J, Zavada J, Pastorekova S;  
XX WPI; 2000-194827/17.  
XX P-PSDB; AAY53228.  
XX Nucleic acid based assay for diagnosing a wide variety of  
XX preneoplastic/neoplastic disease comprises screening for the presence of  
XX abnormal MN gene expression in a vertebrate.  
XX Claim 1; Fig 1; 57pp; English.  
XX The present invention describes a method of screening for  
XX preneoplastic/neoplastic disease. The method comprises: (1) determining  
XX whether abnormal MN gene expression is present in a vertebrate; and (2)  
XX if abnormal MN gene expression is determined to be present in the  
XX vertebrate, determining that the vertebrate has a significant risk of  
XX having preneoplastic/neoplastic disease. The MN gene is an oncogene and  
XX encodes an MN protein (also referred to as MN/CA IX isoenzyme). The MN  
XX protein is a tumor associated carbonic anhydrase isoenzyme. The method  
XX is used for detecting a wide variety of preneoplastic/neoplastic diseases  
XX in a vertebrate, preferably a human. The disease detected is mammary,  
XX bladder, renal, urinary tract, ovarian, uterine, cervical, endometrial,  
XX vaginal, vulvar, prostate, liver, lung, skin, thyroid, pancreatic,  
XX testicular, brain, head and neck, mesodermal, gallbladder, rectal,  
XX duodenal, jejunal, ileal, gastric, pancreatic duct, liver duct, gastric  
XX mucosa, gallbladder epithelium, small intestinal mucosa, colorectal  
XX mucosa, pancreatic duct epithelium or liver duct epithelium  
XX CC preneoplastic/neoplastic disease. AAY16540 to AAY16617 and AAY53228 to  
XX AAY53245 represent sequences used in the exemplification of the present  
XX invention  
SO Sequence 1522 BP; 297 A; 458 C; 453 G; 314 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1522; DB 3; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ACGTCAAGCCGATGCTGCTCCCGTGGCCAGCCCTGAGTCCCTCTGTGATCCCGGCC 60  
DB 1 ACGTCAAGCCGATGCTGCTCCCGTGGCCAGCCCTGAGTCCCTCTGTGATCCCGGCC 60  
QY 61 CTTGCTCAGAGCTCACTGTCGCACTGCTGTCGTCGTCCTTCTGATGCTGTCAT 120  
DB 61 CTTGCTCAGAGCTCACTGTCGCACTGCTGTCGTCGTCCTTCTGATGCTGTCAT 120  
QY 121 CCCGAGAGTTGGCCCGGATGCGAGAGAGTCCCTTGGAGAGAGGCTCTTGGGAA 180  
DB 121 CCCGAGAGTTGGCCCGGATGCGAGAGAGTCCCTTGGAGAGAGGCTCTTGGGAA 180  
QY 181 GATGACCACTGAGGCGAGAGAGATCTGCCAGTGAAGAGAGATTCACCGAGAGAGAT 240  
DB 181 GATGACCACTGAGGCGAGAGAGATCTGCCAGTGAAGAGAGATTCACCGAGAGAGAT 240  
QY 241 CCACCGAGAGAGAGATCTGAGAGAGAGATCTGAGAGAGAGATCTGAGAGAGATCTACT 300

DB 241 CCACCGAGAGAGAGATCTGAGAGAGAGATCTGAGAGAGAGATCTGAGAGAGATCTACT 300  
QY 301 GAAGTTAAGCTTAATATGAAGAGAGAGGCTCCCTGAAGTTAAGATCTACTACTGT 360  
DB 301 GAAGTTAAGCTTAATATGAAGAGAGAGGCTCCCTGAAGTTAAGATCTACTACTGT 360  
QY 361 GAGGCTCTGAGATCTCAAGAACCCAGAAATATGCCACAGAGGACAAAGAGGAGAT 420  
DB 361 GAGGCTCTGAGATCTCAAGAACCCAGAAATATGCCACAGAGGACAAAGAGGAGAT 420  
QY 421 GACGAGATCAATTGAGCGCTATGAGAGGAGACCCGCTGCGCGGAGGTCTCCAGCTGC 480  
DB 421 GACGAGATCAATTGAGCGCTATGAGAGGAGACCCGCTGCGCGGAGGTCTCCAGCTGC 480  
QY 481 GCGGCGGCTTCAGATCCCGGATGATCCGAGCTCCGAGCTGCGCGCTCTGCGGCG 540  
DB 481 GCGGCGGCTTCAGATCCCGGATGATCCGAGCTCCGAGCTGCGCGCTCTGCGGCG 540  
QY 541 CTGCGCCCTCTGGAATCTGAGGCTTCCAGCTCCGCGCTCCAGAACTGCGCTGCGC 600  
DB 541 CTGCGCCCTCTGGAATCTGAGGCTTCCAGCTCCGCGCTCCAGAACTGCGCTGCGC 600  
QY 601 AACATGCGCAGATGTCACATGACCTGCTCTGAGGCTAGAGATGCTGAGTCC 660  
DB 601 AACATGCGCAGATGTCACATGACCTGCTCTGAGGCTAGAGATGCTGAGTCC 660  
QY 661 GGGCGGAGTACCGGCTCTGAGCTGATCTGCACTGGGAGGCTGAGGCTGCGG 720  
DB 661 GGGCGGAGTACCGGCTCTGAGCTGATCTGCACTGGGAGGCTGAGGCTGCGG 720  
QY 721 TGGAGACACATGTCGAGAGGACACCGTTCCCGGAGATCCACGTCGTCACG 780  
DB 721 TGGAGACACATGTCGAGAGGACACCGTTCCCGGAGATCCACGTCGTCACG 780  
QY 781 ACCGCTTTCAGAGTTCAGAGGACCTTGGGAGGCTGAGGAGGCTGAGGCTG 840  
DB 781 ACCGCTTTCAGAGTTCAGAGGACCTTGGGAGGCTGAGGAGGCTGAGGCTG 840  
QY 841 GCGTTTTCGAGAGAGGCGCGGAGAGAAAGTGCCTTATGAGCATGTCGTCG 900  
DB 841 GCGTTTTCGAGAGAGGCGCGGAGAGAAAGTGCCTTATGAGCATGTCGTCG 900  
QY 901 GAGGAATGCTGAGAGAGGCTCAGAGCTCAGAGTCCAGAGCTGACATATGCACTC 960  
DB 901 GAGGAATGCTGAGAGAGGCTCAGAGCTCAGAGTCCAGAGCTGACATATGCACTC 960  
QY 961 CTGCGCTGATTCAGCGCTCACTTCCAAATATGAGGAGTCTGATCAACCGCTCT 1020  
DB 961 CTGCGCTGATTCAGCGCTCACTTCCAAATATGAGGAGTCTGATCAACCGCTCT 1020  
QY 1021 GCCCAGGAGTATCTGAGACTGTGTTTACACAGACAGATGCTGATGACAGCTC 1080  
DB 1021 GCCCAGGAGTATCTGAGACTGTGTTTACACAGACAGATGCTGATGACAGCTC 1080  
QY 1081 CACACCTCTGACACCTCTGAGGAGCTGAGTCTGAGTCTGAGTCTGAGTCTG 1140  
DB 1081 CACACCTCTGACACCTCTGAGGAGCTGAGTCTGAGTCTGAGTCTGAGTCTG 1140  
QY 1141 GCGAGCAGACCTTGAATGAGGAGATGATGAGGCTCTTCCCTGAGAGTGAAGC 1200  
DB 1141 GCGAGCAGACCTTGAATGAGGAGATGATGAGGCTCTTCCCTGAGAGTGAAGC 1200  
QY 1201 AGTCTCTGAGGCTGAGGACAGTCCAGCTGATTCCTGCTGCTGCTGCTGATCTTA 1260  
DB 1201 AGTCTCTGAGGCTGAGGACAGTCCAGCTGATTCCTGCTGCTGCTGCTGATCTTA 1260  
QY 1261 GCGCTGCTTCTGAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
DB 1261 GCGCTGCTTCTGAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
QY 1321 AGGACACACAGAGAGGAGACCAAGAGGAGTGAAGTACCGCCAGCAGAGTACCGAG 1380  
DB 1321 AGGACACACAGAGAGGAGACCAAGAGGAGTGAAGTACCGCCAGCAGAGTACCGAG 1380

QY 1361 ACTGGAGCTAGAGCTGATCTTGGAAATGTGAGAGCCAGGCACTCTGAGG 1440  
DB 1361 ACTGGAGCTAGAGCTGATCTTGGAAATGTGAGAGCCAGGCACTCTGAGG 1440  
QY 1441 GGAGCGGTAACTGCTCTGCTCTGCTCATTAATGCACTTCTTTAACTGCCAAGAAAT 1500  
DB 1441 GGAGCGGTAACTGCTCTGCTCTGCTCATTAATGCACTTCTTTAACTGCCAAGAAAT 1500  
QY 1501 TTTTAAATTAATTAATTTTAAAT 1522  
DB 1501 TTTTAAATTAATTAATTTTAAAT 1522

RESULT 3  
AAA52459  
ID AAA52459 standard; cDNA; 1522 BP.  
XX  
AC AAA52459;  
XX  
DT 25-SEP-2000 (first entry)  
XX  
DE Human MN cDNA.  
XX  
KW MN protein; tumour associated cell adhesion molecule; oncoprotein;  
KW proteoglycan domain; PG domain; carbonic anhydrase; CA domain;  
KW abnormal expression; neoplastic disease; cancer; gene therapy; ss.  
OS Homo sapiens.  
XX  
FT Key Location/Qualifiers  
FT CDS 13..1392  
FT /tag= a  
FT /product= "Human MN protein"

W020024913-A2.  
XX  
PD 04-MAY-2000.  
XX  
PF 22-OCT-1999; 99WO-US024879.  
XX  
PR 23-OCT-1998; 98US-00177776.  
XX  
PR 23-OCT-1998; 98US-00178115.  
XX  
PA (FARB ) BAYER CORP.  
XX  
PA (VIRO-) INST VIROLOGY.  
XX  
PI Zavada J, Pastorekova S, Pastorek J;  
XX  
DR WPI; 2000-350752/30.  
XX  
DR P-PSDB; AAB03005.  
XX  
PT A molecule which specifically binds to a site on MN protein (oncoprotein)  
PT and prevents adhesion of vertebrate cells to the protein, useful for  
PT treating preneoplastic or neoplastic diseases such as cancer.  
XX  
PS Example 1; Fig 1A-C; 154pp; English.  
XX  
CC The invention relates to the inhibition of cell adhesion mediated by the  
CC MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G350  
CC protein). The MN protein is a tumour-associated adhesion molecule which  
CC comprises a proteoglycan-like (PG) domain (AAB03017) which contains the  
CC protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018).  
CC Abnormal expression of the MN protein is associated with tumorigenicity.  
CC The invention encompasses molecules (e.g., proteins and peptides) which  
CC which specifically bind to a site on the MN protein, thereby preventing  
CC adhesion of vertebrate cells to the protein in a cell adhesion assay. It  
CC also encompasses MN proteins or MN protein fragments which can be added  
CC to the extracellular environment to prevent the adhesion of vertebrate  
CC cells to each other. The invention also relates to the identification of  
CC the binding site of the MN protein and to a method of identifying a site  
CC on an MN protein to which cells adhere, comprising testing a series of  
CC overlapping peptides from the protein in a cell adhesion assay. The

CC invention encompasses a vector comprising an expression control sequence  
CC operatively linked to a nucleic acid encoding the variable domains of a  
CC MN-specific antibody, where the domains are separated by a flexible  
CC linker peptide (AAB03035) and the vector inhibits the growth of a  
CC vertebrate preneoplastic or neoplastic cell that abnormally expresses MN  
CC protein. The invention also encompasses a vector comprising a nucleic  
CC acid encoding a cytotoxic protein or peptide operatively linked to the  
CC MN gene promoter, which inhibits the growth of a vertebrate preneoplastic or  
CC neoplastic cell. Also claimed is a repressor complex that binds to the MN  
CC gene promoter (AAA52473). MN proteins and peptides, MN-binding proteins  
CC and peptides, and expression vectors encoding such proteins and peptides  
CC are useful for treating patients with preneoplastic or neoplastic disease  
CC (e.g., cancers) associated with or characterized by abnormal MN  
CC expression. The present sequence represents cDNA encoding the human MN  
CC protein  
XX  
SQ Sequence 1522 BP; 297 A; 458 C; 453 G; 314 T; 0 U; 0 Other;  
XX  
Query Match 100.0%; Score 1522; DB 3; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACACTCAGAGCCGATGAGCTCCCTGTCGCCCAAGCCCTGAGCTCCTCTGTGATCCCGGCC 60  
DB 1 ACACTCAGAGCCGATGAGCTCCCTGTCGCCCAAGCCCTGAGCTCCTCTGTGATCCCGGCC 60  
QY CTTGCTTCAGAGCTCACTGTCGCACTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 120  
DB CTTGCTTCAGAGCTCACTGTCGCACTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 120  
QY 121 CCCCAGAGGTTGCCCCGAGATGCGAGAGATCCCTTGGAGAGAGCTCTTGGGAA 180  
DB 121 CCCCAGAGGTTGCCCCGAGATGCGAGAGATCCCTTGGAGAGAGCTCTTGGGAA 180  
QY 121 CCCCAGAGGTTGCCCCGAGATGCGAGAGATCCCTTGGAGAGAGCTCTTGGGAA 180  
DB 121 CCCCAGAGGTTGCCCCGAGATGCGAGAGATCCCTTGGAGAGAGCTCTTGGGAA 180  
QY 181 GATGATCCCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCAACCAAGAGAGAGAT 240  
DB 181 GATGATCCCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCAACCAAGAGAGAGAT 240  
QY 241 CCACCGGAGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACCT 300  
DB 241 CCACCGGAGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACCT 300  
QY 301 GAAATTAAAGCTTAATCAAGAGAGAGAGCTCTTGAAGTGAAGAGATCTACTACTGTT 360  
DB 301 GAAATTAAAGCTTAATCAAGAGAGAGAGCTCTTGAAGTGAAGAGATCTACTACTGTT 360  
QY 361 GAGGCTCTGAGATCTCAAGAACCCCAATTAATGCCACAGGAGCAAGAGAGAGAT 420  
DB 361 GAGGCTCTGAGATCTCAAGAACCCCAATTAATGCCACAGGAGCAAGAGAGAGAT 420  
QY 421 GACCAAGATCATTTGAGCTATGAGGAGACCCGCTGAGCCCGGAGTGTCCCAAGCTGC 480  
DB 421 GACCAAGATCATTTGAGCTATGAGGAGACCCGCTGAGCCCGGAGTGTCCCAAGCTGC 480  
QY 481 GCGGCGCGCTTCAAGTCCCGGATGATCCCGCCCAAGCTGCGGCTTCTGCGCGGC 540  
DB 481 GCGGCGCGCTTCAAGTCCCGGATGATCCCGCCCAAGCTGCGGCTTCTGCGCGGC 540  
QY 541 CTGGGCCCCCTGAGATCTCTGAGCTTCCAGCTCCGCGCTCCCAAACTGCGCTGCGC 600  
DB 541 CTGGGCCCCCTGAGATCTCTGAGCTTCCAGCTCCGCGCTCCCAAACTGCGCTGCGC 600  
QY 601 AACATATGCGCAAGTGTGCACTGACCTGCTCTCTGAGCTGAGATGCTTGGAGTCC 660  
DB 601 AACATATGCGCAAGTGTGCACTGACCTGCTCTCTGAGCTGAGATGCTTGGAGTCC 660  
QY 661 GAGCGGAGATACCGGAGCTGAGCTGATGCACTGAGGAGGAGTGTGCTGCGGC 720  
DB 661 GAGCGGAGATACCGGAGCTGAGCTGATGCACTGAGGAGGAGTGTGCTGCGGC 720  
QY 721 TCGAGAGCACTGTGAGAGGAGCCAGCTTCTCTGCGGAGATCCAGTGGTTCACTGAGC 780  
DB 721 TCGAGAGCACTGTGAGAGGAGCCAGCTTCTCTGCGGAGATCCAGTGGTTCACTGAGC 780

QY 781 ACCGCTTTGCGAGAGTTGACGAGGCTTGGGGCCCGCGAGGCGCTGGCCCTGTTGGCC 840  
 DB 781 ACCGCTTTGCGAGAGTTGACGAGGCTTGGGGCCCGCGAGGCGCTGGCCCTGTTGGCC 840  
 QY 841 GCGTTTCTGAGAGAGGCGCGAGAGAAACAGTGTCTATGACAGTTGCTGTCTGCTTG 900  
 DB 841 GCGTTTCTGAGAGAGGCGCGAGAGAAACAGTGTCTATGACAGTTGCTGTCTGCTTG 900  
 QY 901 GAAGAAATGCTGAGAGAGGCTCAGAGACTCAGGTCCAGAGACTGAGACATATCTGACCTC 960  
 DB 901 GAAGAAATGCTGAGAGAGGCTCAGAGACTCAGGTCCAGAGACTGAGACATATCTGACCTC 960  
 QY 961 CTGCGCTCTGACTTCAAGCGCTACTTCCAAATATGAGGGGTCTGACTACACCGCCCTGT 1020  
 DB 961 CTGCGCTCTGACTTCAAGCGCTACTTCCAAATATGAGGGGTCTGACTACACCGCCCTGT 1020  
 QY 1021 GCGCAGGCTGATCTGAGACTGTGTTAACAGAGAGTGAATGCTGAGTGTCTAGAGACTC 1080  
 DB 1021 GCGCAGGCTGATCTGAGACTGTGTTAACAGAGAGTGAATGCTGAGTGTCTAGAGACTC 1080  
 QY 1081 CACACCTCTCTGACACCTCTGAGGACCTGCTGACTCTCGGCTCAACAGCTGAACTTCCGA 1140  
 DB 1081 CACACCTCTCTGACACCTCTGAGGACCTGCTGACTCTCGGCTCAACAGCTGAACTTCCGA 1140  
 QY 1141 GCGAGCGACGCTTTGAAATGAGGCGAGTGAATGAGGCTCTCTGCTGAGTGAAGCAGC 1200  
 DB 1141 GCGAGCGACGCTTTGAAATGAGGCGAGTGAATGAGGCTCTCTGCTGAGTGAAGCAGC 1200  
 QY 1201 AGTCTTGCGGCTGCTGAGGAGCTGAGTGAATGCTGCTGCTGCTGCTGCTGAGTGAATGCTG 1260  
 DB 1201 AGTCTTGCGGCTGCTGAGGAGCTGAGTGAATGCTGCTGCTGCTGCTGCTGAGTGAATGCTG 1260  
 QY 1261 GCGCTGCTTTTGGCTCTCTTCTGCTGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
 DB 1261 GCGCTGCTTTTGGCTCTCTTCTGCTGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
 QY 1321 AGGAGACAGAGAGAGGAG 1380  
 DB 1321 AGGAGACAG 1380  
 QY 1381 ACTGAGAGCTAGAGAGCTGAGTCTTGAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 DB 1381 ACTGAGAGCTAGAGAGCTGAGTCTTGAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1441 GAGAGCGGTAAGTCTGCTCTGCTCATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 DB 1441 GAGAGCGGTAAGTCTGCTCTGCTCATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY 1501 TTTTAAATTAATTAATTAAT 1522  
 DB 1501 TTTTAAATTAATTAATTAAT 1522

RESULT 4  
 ABL64779  
 ID ABL64779 standard; DNA; 1552 BP.  
 XX  
 AC ABL64779;  
 XX  
 DT 15-MAY-2002 (first entry)  
 XX  
 DE Lung cancer related gene sequence SEQ ID NO:3116.  
 XX  
 XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 XX stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;  
 XX cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;  
 XX gene; ds.  
 OS Homo sapiens.  
 XX  
 PN MO200194629-A2.  
 XX

PD 13-DEC-2001.  
 XX  
 PF 30-MAY-2001; 2001WO-US010838.  
 XX  
 PR 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-023133P.  
 PR 18-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237112P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 XX (AVAL-) AVALON PHARM.  
 XX  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 XX WPI; 2002-188264/24.  
 DR  
 XX  
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 XX Claim 1, SEQ ID NO 3116; 44pp; English.  
 PS  
 XX The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in (1)  
 CC expression of at least one gene (1) of a signature gene set, where (1)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (1) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,



PR 22-SEP-2000; 2000US-0234503P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237605P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 XX (AVAL-) AVALON PHARM.  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX WPI; 2002-188264/24.  
 XX  
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 PS Claim 1; SEQ ID NO 516; 44pp; English.  
 XX  
 CC The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in Ab161664  
 CC to Ab170110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms  
 CC tumour  
 CC  
 SO Sequence 1552 BP; 302 A; 471 C; 461 G; 318 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 1522; DB 6; Length 1552;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGTCAGCCGATGAGTCTCCCTGTGCCCCAGCCCTTGAGTCTCTGTGATCCCGGCC 60  
 Db 31 AAGTCAGCCGATGAGTCTCCCTGTGCCCCAGCCCTTGAGTCTCTGTGATCCCGGCC 90  
 QY 61 CTTGCTCCAGGCTTCATGTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 Db 91 CTTGCTCCAGGCTTCATGTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 150  
 QY 121 CCCGAGAGTGGCCCGGATGACAGAGAGATTCCTTGGAGAGAGCTCTTGGGGA 180  
 Db 151 CCCGAGAGTGGCCCGGATGACAGAGAGATTCCTTGGAGAGAGCTCTTGGGGA 210  
 QY 181 GATGACCTGCTGGCCGAGAGAGATCTGCTCCAGTGAAGAGATTCACCGAGAGAGAT 240  
 Db 211 GATGACCTGCTGGCCGAGAGAGATCTGCTCCAGTGAAGAGATTCACCGAGAGAGAT 270  
 QY 241 CCACCCGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGATCTACCT 300  
 Db 271 CCACCCGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGATCTACCT 330  
 QY 301 GAAGTTAAGCTTAATGAGAAAGAGAGGCTCCCTGAGTTAGAGATCTACCTACTGTT 360  
 Db 331 GAAGTTAAGCTTAATGAGAAAGAGAGGCTCCCTGAGTTAGAGATCTACCTACTGTT 390  
 QY 361 GAGGCTCTGAGATCTCTCAAGAAAGAGAGATTAAGCCAGAGAGAGAGAGAGAT 420  
 Db 391 GAGGCTCTGAGATCTCTCAAGAAAGAGAGATTAAGCCAGAGAGAGAGAGAGAT 450  
 QY 421 GACCAAGTCAATTGGCGCTATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 Db 451 GACCAAGTCAATTGGCGCTATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 510  
 QY 481 GCGGCGGCTTCCAGTCCCGGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 Db 511 GCGGCGGCTTCCAGTCCCGGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570  
 QY 541 CTGGCGCCCTGGAATCTGAGGCTTCCAGCTCCCGGCTCCAGAACTCGGCTCGCG 600  
 Db 571 CTGGCGCCCTGGAATCTGAGGCTTCCAGCTCCCGGCTCCAGAACTCGGCTCGCG 630  
 QY 601 AACCAATGGCCAGTGTGCACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 Db 631 AACCAATGGCCAGTGTGCACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 690  
 QY 661 GGGCGGAGATCCCGGCTCTGCAAGCTGCACTGCACTGCACTGCACTGCACTGCACTG 720  
 Db 691 GGGCGGAGATCCCGGCTCTGCAAGCTGCACTGCACTGCACTGCACTGCACTGCACTG 750  
 QY 721 TCGAGGACACTGTGGAAGGCAAGCTTCCCTGCGAGATCCAGTGTGCTACCTGACG 780  
 Db 751 TCGAGGACACTGTGGAAGGCAAGCTTCCCTGCGAGATCCAGTGTGCTACCTGACG 810  
 QY 781 ACCGCTTGTGCAAGTGTGAGAGGCTTGGGCGCCCGAGAGAGCTGCGCTGTGGCC 840  
 Db 811 ACCGCTTGTGCAAGTGTGAGAGGCTTGGGCGCCCGAGAGAGCTGCGCTGTGGCC 870  
 QY 841 GCTTTTGTGAGAGGCGCCGAGAAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 Db 871 GCTTTTGTGAGAGGCGCCGAGAAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 930  
 QY 901 GAAAGAAATGCTGAGAGAGGCTGAGAGCTCAGGTCACAGAGCTGACATATGCACTC 960  
 Db 931 GAAAGAAATGCTGAGAGAGGCTGAGAGCTCAGGTCACAGAGCTGACATATGCACTC 990  
 QY 961 CTGCTCTGACTTGAAGCGCTGACTTCCATATGAGAGAGGCTGCTGACTAATGCACTG 1020  
 Db 991 CTGCTCTGACTTGAAGCGCTGACTTCCATATGAGAGAGGCTGCTGACTAATGCACTG 1050  
 QY 1021 GCCCAGGATGATCTGAGTGTGTTAACAAGACAGTGTGCTGAGTGTGAGTGTGAGTGT 1080  
 Db 1051 GCCCAGGATGATCTGAGTGTGTTAACAAGACAGTGTGCTGAGTGTGAGTGTGAGTGT 1110  
 QY 1081 CACACCTCTGTGACACCTGTGGGAGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1140

```

Db      1111  CACACCTCTGTGACACCTGTGGGACCTGTGACTCTGCGCTACAGCTGAATCCCA 1170
Qy      1141  GCGAGCGACCTTTGATGAGCGAGTGAATGAGGCTCTTCCCTGCTGAGTGAACAGC 1200
Db      1171  GCGAGCGACCTTTGATGAGCGAGTGAATGAGGCTCTTCCCTGCTGAGTGAACAGC 1230
Qy      1201  AGTCTCGGGGCTGCTGAGCAGTCAAGTGAATCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db      1231  AGTCTCGGGGCTGCTGAGCAGTCAAGTGAATCTGCTGCTGCTGCTGCTGCTGCTG 1290
Qy      1261  GCGCTGCTTTTGGCTCTTCTTCTGTCACGACGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db      1291  GCGCTGCTTTTGGCTCTTCTTCTGTCACGACGCTGCTGCTGCTGCTGCTGCTGCTG 1350
Qy      1321  AGGCGACGACAGAGGGGAGCAACAGAGGGGCTGAGCTACCGCCAGCAGAGGAGCGAG 1380
Db      1351  AGGCGACGACAGAGGGGAGCAACAGAGGGGCTGAGCTACCGCCAGCAGAGGAGCGAG 1410
Qy      1381  ACTGAGGCTTGAAGGCTGATCTTGAGAAATGTGAGAGCCAGCCAGGCAATCTGAGGG 1440
Db      1411  ACTGAGGCTTGAAGGCTGATCTTGAGAAATGTGAGAGCCAGCCAGGCAATCTGAGGG 1470
Qy      1441  GGAGCGGCTTGAAGGCTGATCTTGAGAAATGTGAGAGCCAGCCAGGCAATCTGAGGG 1500
Db      1471  GGAGCGGCTTGAAGGCTGATCTTGAGAAATGTGAGAGCCAGCCAGGCAATCTGAGGG 1530
Qy      1501  TTTTAAATTAATATTATTAT 1522
Db      1531  TTTTAAATTAATATTATTAT 1552

RESULT 6
ABL65416
ID      ABL65416 standard; DNA; 1552 BP.
XX
XX      ABL65416;
XX
XX      15-MAY-2002 (first entry)
XX
DE      Lung cancer related gene sequence SEQ ID NO:3753.
XX
XX      Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX      stomach; lung; prostate; pancreas; carcinoma; antitumour; carcinos;
XX      cytostatic; gene therapy; anti-neoplastic; Wilms tumour; adenocarcinoma;
XX      gene; de.
XX
XX      Homo sapiens.
XX
XX      WO200194629-A2.
XX
XX      13-DEC-2001.
XX
PF      30-MAY-2001; 2001WO-US010838.
XX
XX      05-JUN-2000; 2000US-0203473P.
XX      05-JUN-2000; 2000US-0203531P.
XX      18-SEP-2000; 2000US-0233133P.
XX      18-SEP-2000; 2000US-0233617P.
XX      20-SEP-2000; 2000US-0234009P.
XX      20-SEP-2000; 2000US-0234034P.
XX      20-SEP-2000; 2000US-0234052P.
XX      22-SEP-2000; 2000US-0234509P.
XX      22-SEP-2000; 2000US-0234567P.
XX      25-SEP-2000; 2000US-0234923P.
XX      25-SEP-2000; 2000US-0234924P.
XX      25-SEP-2000; 2000US-0235077P.
XX      25-SEP-2000; 2000US-0235082P.
XX      25-SEP-2000; 2000US-0235134P.
XX      25-SEP-2000; 2000US-0235280P.
XX      26-SEP-2000; 2000US-0235637P.
XX      26-SEP-2000; 2000US-0235638P.
XX      27-SEP-2000; 2000US-0235711P.

```

```

PR      27-SEP-2000; 2000US-0235720P.
PR      27-SEP-2000; 2000US-0235840P.
PR      27-SEP-2000; 2000US-0235863P.
PR      28-SEP-2000; 2000US-0236028P.
PR      28-SEP-2000; 2000US-0236032P.
PR      28-SEP-2000; 2000US-0236033P.
PR      28-SEP-2000; 2000US-0236034P.
PR      28-SEP-2000; 2000US-0236109P.
PR      28-SEP-2000; 2000US-0236111P.
PR      29-SEP-2000; 2000US-0236842P.
PR      29-SEP-2000; 2000US-0236891P.
PR      02-OCT-2000; 2000US-0237172P.
PR      02-OCT-2000; 2000US-0237173P.
PR      02-OCT-2000; 2000US-0237278P.
PR      02-OCT-2000; 2000US-0237294P.
PR      02-OCT-2000; 2000US-0237295P.
PR      02-OCT-2000; 2000US-0237316P.
PR      03-OCT-2000; 2000US-0237425P.
PR      03-OCT-2000; 2000US-0237598P.
PR      03-OCT-2000; 2000US-0237604P.
PR      03-OCT-2000; 2000US-0237606P.
PR      03-OCT-2000; 2000US-0237608P.
PR      01-NOV-2000; 2000US-0244867P.
PR      01-NOV-2000; 2000US-0245084P.
XX
XX      (AVAL-) AVALON PHARM.
XX
XX      Young PE, Augustus M, Carter KC, Ebnor R, Endress G, Horrigan S;
XX      Soppet DR, Weaver Z;
XX      WPI, 2002-188264/24.
XX
PT      Screening for anti-neoplastic agent involves exposing cells to a chemical
PT      agent to be tested for anti-neoplastic activity, and determining a change
PT      in expression of a gene of a signature gene set.
XX
PS      Claim 1; SEQ ID NO 3753; 44p; English.
XX
XX      The present invention describes a method (M1) for screening for an anti-
XX      neoplastic agent. The method involves exposing cells to a chemical agent
XX      to be tested for anti-neoplastic activity, determining a change in
XX      expression of at least one gene (I) of a signature gene set, where (I)
XX      comprises a sequence (S) selected from 9447 sequences (given in ABL6164
XX      to ABL70110), or is at least 95% identical to (S), where a change in
XX      expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX      activity and can be used in gene therapy. M1 can be used for screening an
XX      anti-neoplastic agent, and can be used for producing a product which is
XX      the data collected with respect to the anti-neoplastic agent as a result
XX      of M1, and the data is sufficient to convey the chemical structure and/or
XX      properties of the agent. M1 can be used in the treatment of cancer such
XX      as colon, breast, stomach, lung, thyroid, oesophagal, ovarian, kidney,
XX      prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
XX      cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
XX      cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms's
XX      tumour.
XX
XX      Sequence 1552 BP; 302 A; 471 C; 461 G; 318 T; 0 U; 0 Other;
XX
XX      Query Match      100.0%; Score 1522; DB 6; Length 1552;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1  ACAATGACCGGATGCTCCGCTGCCCGACGCGCTGCTCCCTGTTGATCCCGGCC 60
Db      31  ACAATGACCGGATGCTCCGCTGCCCGACGCGCTGCTCCCTGTTGATCCCGGCC 90
Qy      61  CCTGCTCAGGCTCACTGTGCAACTGCTGCTGCACTGCTGCTGCTGCTGCTGCTGCT 120
Db      91  CCTGCTCAGGCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 150
Qy      121  CCCCAGAGGTTGCCCGGATGCGAGAGATTCCCTTTGGAGAGAGCTCTTCTGGGAGA 180
Db      151  CCCCAGAGGTTGCCCGGATGCGAGAGATTCCCTTTGGAGAGAGCTCTTCTGGGAGA 210

```

QY 181 GATGACCACTGAGGAGGATCTGCCAGTGAAGATTCACCCAGAGAGAGAT 240  
DB 211 GATGACCACTGAGGAGGATCTGCCAGTGAAGATTCACCCAGAGAGAGAT 270  
QY 241 CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGATCTACT 300  
DB 271 CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGATCTACT 330  
QY 301 GAAGTTAGCTTAATCAAGAGAGAGAGATCTACCTGAGAGAGATCTACTACT 360  
DB 331 GAAGTTAGCTTAATCAAGAGAGAGAGATCTACCTGAGAGAGATCTACTACT 390  
QY 361 GAGGCTCTGAGAGATCTCAAGAACCCCAATTAATGCCCAAGGAGCAAGAGAGAT 420  
DB 391 GAGGCTCTGAGAGATCTCAAGAACCCCAATTAATGCCCAAGGAGCAAGAGAGAT 450  
QY 421 GACCAAGTCAATGGGCTATGAGAGGAGCCGCGCTGAGCGCGCTGAGCGCTGAG 480  
DB 451 GACCAAGTCAATGGGCTATGAGAGGAGCCGCGCTGAGCGCGCTGAGCGCTGAG 510  
QY 481 GCGGCGCGCTTCCAGTCCCGAGTATCCGCGCGCTGAGCGCGCTTCCAGTCCCG 540  
DB 511 GCGGCGCGCTTCCAGTCCCGAGTATCCGCGCGCTGAGCGCGCTTCCAGTCCCG 570  
QY 541 CTGCGCGCGCTTCCAGTCCCGAGTATCCGCGCGCTGAGCGCGCTTCCAGTCCCG 600  
DB 571 CTGCGCGCGCTTCCAGTCCCGAGTATCCGCGCGCTGAGCGCGCTTCCAGTCCCG 630  
QY 601 AACATGAGCAGAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 660  
DB 631 AACATGAGCAGAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 690  
QY 661 GGGCGGAGTACCGGAGTCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 720  
DB 691 GGGCGGAGTACCGGAGTCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 750  
QY 721 TCGAGACACACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
DB 751 TCGAGACACACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 810  
QY 781 ACCGCTTTTCCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
DB 811 ACCGCTTTTCCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 870  
QY 841 GCCTTTCTGAG 900  
DB 871 GCCTTTCTGAG 930  
QY 901 GAAGAAATGCTGAG 960  
DB 931 GAAGAAATGCTGAG 990  
QY 961 CTGCGCTTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 991 CTGCGCTTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1050  
QY 1021 GCCCAGAGTGTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 1051 GCCCAGAGTGTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1110  
QY 1081 CACACCTCTGAG 1140  
DB 1111 CACACCTCTGAG 1170  
QY 1141 GCGAGCAGAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
DB 1171 GCGAGCAGAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1230  
QY 1201 AGTCTCTGAG 1260  
DB 1231 AGTCTCTGAG 1290

QY 1261 GCCCTGTTTGGCTTCTTTTGTCTGACAGCGCTGCTTCTTGTGAGATGAGA 1320  
DB 1291 GCCCTGTTTGGCTTCTTTTGTCTGACAGCGCTGCTTCTTGTGAGATGAGA 1350  
QY 1321 AGGCAGCAG 1380  
DB 1351 AGGCAGCAG 1410  
QY 1381 ACTGAGCTTGAAGCTGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1411 ACTGAGCTTGAAGCTGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAGAG 1470  
QY 1441 GAGCGGCTTGAAGCTGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAG 1500  
DB 1471 GAGCGGCTTGAAGCTGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAG 1530  
QY 1501 TTTTAAATTAATTTATAT 1552  
DB 1531 TTTTAAATTAATTTATAT 1552

## RESULT 7

ABL68346  
ID ABL68346 standard; DNA; 1552 BP.

XX ABL68346;

DT 15-MAY-2002 (first entry)

XX Kidney cancer related gene sequence SEQ ID NO:6893.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

KM cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;

OS gene; ds.

XX Homo sapiens.

XX WO200194629-A2.

XX 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US010838.

XX 05-JUN-2000; 2000US-0209473P.

PR 05-JUN-2000; 2000US-0209531P.

PR 18-SEP-2000; 2000US-0233133P.

PR 18-SEP-2000; 2000US-0233617P.

PR 20-SEP-2000; 2000US-0234009P.

PR 20-SEP-2000; 2000US-0234034P.

PR 20-SEP-2000; 2000US-0234052P.

PR 22-SEP-2000; 2000US-0234509P.

PR 22-SEP-2000; 2000US-0234567P.

PR 25-SEP-2000; 2000US-0234923P.

PR 25-SEP-2000; 2000US-0234924P.

PR 25-SEP-2000; 2000US-0235077P.

PR 25-SEP-2000; 2000US-0235082P.

PR 25-SEP-2000; 2000US-0235134P.

PR 25-SEP-2000; 2000US-0235280P.

PR 26-SEP-2000; 2000US-0235637P.

PR 26-SEP-2000; 2000US-0235711P.

PR 27-SEP-2000; 2000US-0235720P.

PR 27-SEP-2000; 2000US-0235840P.

PR 27-SEP-2000; 2000US-0235863P.

PR 28-SEP-2000; 2000US-0236028P.

PR 28-SEP-2000; 2000US-0236032P.

PR 28-SEP-2000; 2000US-0236033P.

PR 28-SEP-2000; 2000US-0236034P.

PR 28-SEP-2000; 2000US-0236109P.

PR 28-SEP-2000; 2000US-0236111P.

PR 29-SEP-2000; 2000US-0236842P.

PR 29-SEP-2000; 2000US-0236891P.



PR 02-OCT-2000; 2000US-0237172P.  
PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237284P.  
PR 02-OCT-2000; 2000US-0237295P.  
PR 02-OCT-2000; 2000US-0237316P.  
PR 02-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.  
PR 01-NOV-2000; 2000US-0245084P.

(AVAL-) AVALON PHARM.

PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.

PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.

Claim 1; SEQ ID NO 6683; 44bp; English.

XX The present invention describes a method (M1) for screening for an anti-  
XX neoplastic agent. The method involves exposing cells to a chemical agent  
XX to be tested for anti-neoplastic activity, determining a change in  
XX expression of at least one gene (1) of a signature gene set, where (1)  
XX comprises a sequence (S) selected from 8447 sequences (given in ABL1564  
XX to ABL70110), or is at least 95% identical to (S), where a change in  
XX expression is indicative of anti-neoplastic activity. (1) has cytosolic  
XX activity and can be used in gene therapy. M1 can be used for screening an  
XX anti-neoplastic agent, and can be used for producing a product which is  
XX the data collected with respect to the anti-neoplastic agent as a result  
XX of M1, and the data is sufficient to convey the chemical structure and/or  
XX properties of the agent. M1 can be used in the treatment of cancer such  
XX as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
XX prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
XX cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
XX cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms  
XX tumour

XX Sequence 1552 BP; 302 A; 471 C; 461 G; 318 T; 0 U; 0 Other;

Query Match 100.0%; Score 1552; DB 6; Length 1552;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTCAGCCGATGCTCCCTGTGCCCAAGCCCTGCTCCTCTGTGATCCGAGCC 60  
DB 31 AAGATCAGCCGATGCTCCCTGTGCCCAAGCCCTGCTCCTCTGTGATCCGAGCC 90  
QY 61 CCTGCTCAGGCTCTCACTGTGCACTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 120  
DB 91 CCTGCTCAGGCTCTCACTGTGCACTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 150  
QY 121 CCCGAGAGTTGCCCGGATGTCAGAGAGATCCCTTGGAGAGAGCTCTTTCGAGAA 180  
DB 151 CCCGAGAGTTGCCCGGATGTCAGAGAGATCCCTTGGAGAGAGCTCTTTCGAGAA 210  
QY 181 GATGACCACTGCGGAGAGAGATCTGCCAGTGAAGAGATTACCCAGAGAGAGAT 240  
DB 211 GATGACCACTGCGGAGAGAGATCTGCCAGTGAAGAGATTACCCAGAGAGAGAT 270  
QY 241 CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGAGATCTACCT 300  
DB 271 CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGAGATCTACCT 330  
QY 301 GAAGTTAAGCTTAATCAGAGAGAGAGCTCCTGAAAGTTAGAGATCTACTCTTT 360

DB 331 GAAGTTAAGCTTAATCAGAGAGAGAGCTCCTGAAAGTTAGAGATCTACTCTTT 390  
QY 361 GAGGCTCTGAGATCTCTAAGAACCCCAAGTAATATGCCACAGGACCAAGAGAGAT 420  
DB 391 GAGGCTCTGAGATCTCTAAGAACCCCAAGTAATATGCCACAGGACCAAGAGAGAT 450  
QY 421 GACCAAGATATTGGGCTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480  
DB 451 GACCAAGATATTGGGCTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 510  
QY 481 GCGGAGAGATGAT 540  
DB 511 GCGGAGAGATGAT 570  
QY 541 CTGAGGAGAGATGAT 600  
DB 571 CTGAGGAGAGATGAT 630  
QY 601 AACAAATGAT 660  
DB 631 AACAAATGAT 690  
QY 661 GAGGAT 720  
DB 691 GAGGAT 750  
QY 721 TCGAT 780  
DB 751 TCGAT 810  
QY 781 ACCGCTTTTCCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 840  
DB 811 ACCGCTTTTCCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 870  
QY 841 GCTTTTGTGAT 900  
DB 871 GCTTTTGTGAT 930  
QY 901 GAT 960  
DB 931 GAT 990  
QY 961 CTGAGGAT 1020  
DB 991 CTGAGGAT 1050  
QY 1021 GCCGAGGAT 1080  
DB 1051 GCCGAGGAT 1110  
QY 1081 CACAGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140  
DB 1111 CACAGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1170  
QY 1141 GCGAT 1200  
DB 1171 GCGAT 1230  
QY 1201 AGTCCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1260  
DB 1231 AGTCCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1290  
QY 1261 GCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1320  
DB 1291 GCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1350  
QY 1321 AGGAT 1380  
DB 1351 AGGAT 1410  
QY 1381 ACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1440  
DB 1411 ACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1470

QY	QY	QY	QY
1441	1471	1501	1531
GGAGCCGGTAATGCTCGTCCTCATATATGCACTTCCTTTAATGCGCAAGAAATT	GGAGCCGGTAATGCTCGTCCTCATATATGCACTTCCTTTAATGCGCAAGAAATT	TTTTAAATTAATATTTTAAAT	TTTTAAATTAATATTTTAAAT
1500	1530	1552	1552

## RESULT 8

ID ACC72730 standard; cDNA; 1552 BP.

AC	ACC72730;
XX	
DT	09-JUL-2003 (first entry)

Human cancer related protein encoding cDNA seq ID NO:69.

Human: cancer; diagnosis; screening; modulator; leukaemia; ischaemia.

heart disease; atherosclerosis; gene; ss

Homo sapiens.

PN WO2003025138-A2

PD 27-MAR-2003.

XX  
PF 17-SEP-2002: 2002WO-US029560.

XX 17-SEP-2001, 2001TTS-0323469P  
DB

PR 20-SEP-2001; 2001US-0323887P.

PR 08-FEB-2002; 2002US-0355145P

PR 12-APR-2002; 2002US-0372246P

XX  
DA (ECSB-) EOS BIOTECHNOLOGY INC[illegible]

PI Zlotnik A;

WPI; 2003-354600/33.

DR F-PSDB; ABKJ0000.  
XX

New genes that are up-regulated in cancer cells are potential markers for diagnosing and

PT therapeutic targets for screening

PS Claim 8; Page 662; 767pp; En

CC The present invention describes

regulates in specific cancer

acute lymphocytic leukemia).

AB55882 to ABR88709. Also described: (1) determining the presence or absence of a pathological cell in a patient; (2) an expression vector comprising a nucleic acid molecule described above; (3) a host cell comprising the vector; (4) an isolated polypeptide, which is encoded by the nucleic acid; (5) an antibody that specifically binds the polypeptide of (4); (6) specifically targeting a compound to a pathological cell in a patient by administering to the patient the antibody above; and (7) a drug screening assay. The nucleic acid is useful as diagnostic markers or therapeutic targets. In particular, the nucleic acid is useful for diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow, bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary, pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases, atherosclerosis and endometriosis. The nucleic acid is also useful in drug screening, particularly for identifying agents for treating these pathologies

SQ Sequence 1552 BP; 302 A; 471 C; 461 G; 318 T; 0 U; 0 Other

Query Match	100.0%;	Score 1522;	DB 7;	Length 1552;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1522; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0.

QY	1	ACAGTCACGGCAATGAGTCTCCCTGTCGCCACAGCCCTGGTCCCTCTGTATGATCCGAGC	60
Db	31	ACAGTCACGGCAATGAGTCTCCCTGTCGCCACAGCCCTGGTCCCTCTGTATGATCCGAGC	90
QY	61	CCTGCTCCAGGCTCACTGTGTCAACTGCTGTCTGTCACTGTCTTCTGTATGCTGTCCAT	120
Db	91	CCTGCTCCAGGCTCACTGTGTCAACTGCTGTCTGTCTGTCACTGTCTTCTGTATGCTGTCCAT	150
QY	121	CCCCAGAGTGTGCCCCCGGATGTCAAGAGAGATTCGCCCTTGGAGAGAGGCTCTTGGGAAA	180
Db	151	CCCCAGAGTGTGCCCCCGGATGTCAAGAGAGATTCGCCCTTGGAGAGAGGCTCTTGGGAAA	210
QY	181	GATGACCCCACTGGGCTGAGAGAGATCTGTGCCAGTGAAGAGATTCACCCAGAGAGAGAT	240
Db	211	GATGACCCCACTGGGCTGAGAGAGATCTGTGCCAGTGAAGAGATTCACCCAGAGAGAGAT	270
QY	241	CCACCCCGAGAGAGAGATCTACCTGTGAGAGAGAGATCTACCTGTGAGAGAGAGATCTACCT	300
Db	271	CCACCCCGAGAGAGAGATCTACCTGTGAGAGAGAGATCTACCTGTGAGAGAGAGATCTACCT	330
QY	301	GAAGTTAAGCCTAAATCAGAAAGAGAGGCTCCCTGAATTGAGAGATCTACCTGTT	360
Db	331	GAAGTTAAGCCTAAATCAGAAAGAGAGGCTCCCTGAATTGAGAGATCTACCTGTT	390
QY	361	GAGGCTCCTGAGATCTCAAGAAACCCCAATATATGCCCAAGAGCAAGAAAGGGAT	420
Db	391	GAGGCTCCTGAGATCTCAAGAAACCCCAATATATGCCCAAGAGCAAGAAAGGGAT	450
QY	421	GACCAAGTCATTTGGCGCTATGAGAGCGACCCGCCCTGGCCCGGAGTGTCCCAAGCTGC	480
Db	451	GACCAAGTCATTTGGCGCTATGAGAGCGACCCGCCCTGGCCCGGAGTGTCCCAAGCTGC	510
QY	481	GGGGGCGGCTTCACTCCCGGTGGATATCCGCCCCCAAGTGGCGGCTTCTGCCCCGGGC	540
Db	511	GGGGGCGGCTTCACTCCCGGTGGATATCCGCCCCCAAGTGGCGGCTTCTGCCCCGGGC	570
QY	541	CTGGCGCCCTTGAAGCTCTGGGCTTCAAGCTCCGCCCGCTCCCAAGATCTGGCTTGGGC	600
Db	571	CTGGCGCCCTTGAAGCTCTGGGCTTCAAGCTCCGCCCGCTCCCAAGATCTGGCTTGGGC	630
QY	601	AACAAATGSCACAGTGTCAACTGACCCCTGCTCTGGGCTTAAGATGGCTTGGGTC	660
Db	631	AACAAATGSCACAGTGTCAACTGACCCCTGCTCTGGGCTTAAGATGGCTTGGGTC	690
QY	661	GGGCGGGAATACCGGGCTCTGACGTGATCTGTGACTGGGGGGCTGCAAGTGTGCGGAGC	720
Db	691	GGGCGGGAATACCGGGCTCTGACGTGATCTGTGACTGGGGGGCTGCAAGTGTGCGGAGC	750
QY	721	TGAGACACACTGTGGAAGGCCACCGTTTCCCTGCGAGATCCACGTGTTCACTCAACG	780
Db	751	TGAGACACACTGTGGAAGGCCACCGTTTCCCTGCGAGATCCACGTGTTCACTCAACG	810
QY	781	ACGGCTTTGGCAGAGTTGACGAGGCTTGGGGGCGCCCGGAGAGGCTGTGGCTGTGGCC	840
Db	811	ACGGCTTTGGCAGAGTTGACGAGGCTTGGGGGCGCCCGGAGAGGCTGTGGCTGTGGCC	870
QY	841	GCCCTTTCTGAGAGAGGCCCGGAGAAAAACGTGCTTATGACAGTTTGTCTGTGCTTG	900
Db	871	GCCCTTTCTGAGAGAGGCCCGGAGAAAAACGTGCTTATGACAGTTTGTCTGTGCTTG	930
QY	901	GAAAGAAATGCTGAGGAAGGCTCAGAGACTAGGCTCCAGAGACTGAGCAATATCTGCACTC	960
Db	931	GAAAGAAATGCTGAGGAAGGCTCAGAGACTAGGCTCCAGAGACTGAGCAATATCTGCACTC	990
QY	961	CTGCGCTCTGACTTACGCGCTACTTCCAAATATAGAGGGTCTGTGACTACACGCGCTGT	1020
Db	991	CTGCGCTCTGACTTACGCGCTACTTCCAAATATAGAGGGTCTGTGACTACACGCGCTGT	1050



```

Db      691 GGGCGGAGTACCGGCTCTGACGTGCACTGCACTGCGGGGGCTGCGAGTCTGCTCCGCGC 750
Qy      721 TCGAGACACTGTGAGAGGACCAAGCTTCCCTGCGAGATCCAGTGTTCACCTAGC 780
Db      751 TCGAGACACTGTGAGAGGACCAAGCTTCCCTGCGAGATCCAGTGTTCACCTAGC 810
Qy      781 ACCGCTTTGCCAGAGTACAGAGGCTTGGGGCGCCCGGAGGCTGCGGTGTGGCC 840
Db      811 ACCGCTTTGCCAGAGTACAGAGGCTTGGGGCGCCCGGAGGCTGCGGTGTGGCC 870
Qy      841 GCCTTCTGAGAGGAGGCGCGAGAGAAACAGTGCCTATGACAGTGTCTGTCTGCTG 900
Db      871 GCCTTCTGAGAGGAGGCGCGAGAGAAACAGTGCCTATGACAGTGTCTGTCTGCTG 930
Qy      901 GAAGAAATCGCTGAGAGAGGCTCAGAGTCAAGGCTCCAGAGCTGACATATCTGCACTC 960
Db      931 GAAGAAATCGCTGAGAGAGGCTCAGAGTCAAGGCTCCAGAGCTGACATATCTGCACTC 990
Qy      961 CTGCGCTCTGACTTTCAGCCGCTACTTCCAAATATGAGGGGTCTCTGACTACACCGCCCTGT 1020
Db      991 CTGCGCTCTGACTTTCAGCCGCTACTTCCAAATATGAGGGGTCTCTGACTACACCGCCCTGT 1050
Qy      1021 GCCCAGGAGTGTCACTGCACTGTGTAAACAGACAGTATGCTGATGCTAAGAGCTC 1080
Db      1051 GCCCAGGAGTGTCACTGCACTGTGTAAACAGACAGTATGCTGATGCTAAGAGAGCTC 1110
Qy      1081 CAACACCTCTCTGACACCTCTGAGGAGCTGTGTACTCTGCGCTCAAGCTGAATCTCCGA 1140
Db      1111 CAACACCTCTCTGACACCTCTGAGGAGCTGTGTACTCTGCGCTCAAGCTGAATCTCCGA 1170
Qy      1141 GCGAGCAGGCTTTGAATGAGGAGATGATGAGGCTCTTCCCTGCTGAGTGAAGCAGC 1200
Db      1171 GCGAGCAGGCTTTGAATGAGGAGATGATGAGGCTCTTCCCTGCTGAGTGAAGCAGC 1230
Qy      1201 AGTCTCTGAGGCTGTGAGCCAGTCCAGTGAATTCCTGCTGCTGTGTGACATCTTA 1260
Db      1231 AGTCTCTGAGGCTGTGAGCCAGTCCAGTGAATTCCTGCTGCTGTGTGACATCTTA 1290
Qy      1261 GCCTGCTTTTGGCTCTCTTTTGTGTGACACAGCGTGGGCTCTGTGTGCAAGTGA 1320
Db      1291 GCCTGCTTTTGGCTCTCTTTTGTGTGACACAGCGTGGGCTCTGTGTGCAAGTGA 1350
Qy      1321 AGCAGCAGACAGAGGGGAAACAAAGGGGGTGTGAGTACCGCCACAGAGATGCCAG 1380
Db      1351 AGCAGCAGACAGAGGGGAAACAAAGGGGGTGTGAGTACCGCCACAGAGATGCCAG 1410
Qy      1381 ACTGAGCCTTGAAGGCTGTGATCTTGAAGATGTGAAGAACCCAGCAGAGGATCTGAGG 1440
Db      1411 ACTGAGCCTTGAAGGCTGTGATCTTGAAGATGTGAAGAACCCAGCAGAGGATCTGAGG 1470
Qy      1441 GGAAGCGGTACTGCTCTGCTGCTCATTAATGCACTCTTTAACTGCAAGAAATT 1500
Db      1471 GGAAGCGGTACTGCTCTGCTGCTCATTAATGCACTCTTTAACTGCAAGAAATT 1530
Qy      1501 TTTTAAATTAATATTATTAAT 1522
Db      1531 TTTTAAATTAATATTATTAAT 1552

```

## RESULT 10

ABX76385 ID ABX76385 standard, DNA, 1552 BP.

XX ABX76385;

XX 02-APR-2003 (first entry)

XX Lung cancer-associated polynucleotide #249.

XX Lung cancer-associated polynucleotide; gene; dr. cytostatic; emphysema;  
 KW anti-inflammatory; antineoplastic; non-small cell lung cancer; atelectasis;  
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchiolitis;

```

KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX Unidentified.
XX MO200286443-A2.
XX 31-OCT-2002.
XX 18-APR-2002; 2002WO-US012476.
XX 18-APR-2001; 2001US-0284770P.
XX 10-MAY-2001; 2001US-0290492P.
XX 09-NOV-2001; 2001US-0338245P.
XX 13-NOV-2001; 2001US-0350666P.
XX 29-NOV-2001; 2001US-0334370P.
XX 12-APR-2002; 2002US-0372246P.
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX Aziz N, Murray R;
XX WPI; 2003-093161/08.
XX P-PSDB; ABUS6656.
XX Detecting a lung cancer-associated transcript in a cell from a patient
XX for treating lung cancer, by contacting a biological sample from the
XX patient with a polynucleotide that exhibits increased or decreased
XX expression in lung cancer.
XX Claim 22; Page 379-380; 453p; English.
XX The invention relates to a method for detecting a lung cancer-associated
XX transcript in a cell from a patient, comprising contacting a biological
XX sample from the patient with a polynucleotide that selectively hybridises
XX to a sequence that is at least 80 % identical to a gene that exhibits
XX increased or decreased expression in lung cancer samples. Lung cancer-
XX associated polynucleotides and polypeptides are used for identifying a
XX compound that modulates a lung cancer-associated polypeptide, for
XX inhibiting proliferation of a lung cancer-associated cell to treat lung
XX cancer in a patient and for treating a mammal having lung cancer by
XX administering a modulatory compound identified. The methods are useful
XX for treating lung cancer, such as small cell lung cancer, non-small cell
XX lung cancer or other benign or precancerous lesions, e.g. atelectasis,
XX emphysema, bronchiolitis, chronic obstructive pulmonary disease, fibrosis,
XX hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
XX bronchiectasis. The genes, polynucleotides and polypeptides are useful
XX for diagnostic purposes and as targets for screening for therapeutic
XX compounds that modulate lung cancer, such as antibodies. Sequences
XX CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
XX invention
XX SO Sequence 1552 BP; 302 A; 471 C; 461 G; 318 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 1522; DB 7; Length 1552;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACAGTCAGCCGATGCTCCCTGTGCCCCGACCCCTGCTGCTCTGTGATCCGCGC 60
Db 31 ACAGTCAGCCGATGCTCCCTGTGCCCCGACCCCTGCTGCTCTGTGATCCGCGC 90
Qy 61 CCTGCTCCAGGCTCTACTGTGTGCACTGTGTGTGCTGCTGCTTGTGAGCCGTGCAT 120
Db 91 CCTGCTCCAGGCTCTACTGTGTGCACTGTGTGTGCTGCTGCTTGTGAGCCGTGCAT 150
Qy 121 CCCCAAGAGTTGCCCGGATCAGAGAGATTCCTCCCTTGGAGAGAGCTCTTGTGGGAA 180
Db 151 CCCCAAGAGTTGCCCGGATCAGAGAGATTCCTCCCTTGGAGAGAGCTCTTGTGGGAA 210
Qy 181 GATGACCACTGGGAGAGAGATTCGCCAGTGAAGATTTACCCCAAGAGAGAT 240
Db 211 GATGACCACTGGGAGAGAGATTCGCCAGTGAAGATTTACCCCAAGAGAGAT 270

```

QY 241 CCACCCGAGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGATCTAACCT 300  
DB 271 CCACCCGAGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGATCTAACCT 330  
QY 301 GAAAGTTAAAGCCTAAATCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 331 GAAAGTTAAAGCCTAAATCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 390  
QY 361 GAGAGCTCTGAG 420  
DB 391 GAGAGCTCTGAG 450  
QY 421 GACAGAGATCATTTGGAG 480  
DB 451 GACAGAGATCATTTGGAG 510  
QY 481 GGGGGCCGCTTCAATCCCCGGTGATATCCGCCCCAGCTCCGCGCTTCTACCGGGCC 540  
DB 511 GGGGGCCGCTTCAATCCCCGGTGATATCCGCCCCAGCTCCGCGCTTCTACCGGGCC 570  
QY 541 CTGCGCCGCTTGAAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
DB 571 CTGCGCCGCTTGAAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630  
QY 601 AACAAATGACAG 660  
DB 631 AACAAATGACAG 690  
QY 661 GGGGGGAG 720  
DB 691 GGGGGGAG 750  
QY 721 TCGAGACACTGTGAG 780  
DB 751 TCGAGACACTGTGAG 810  
QY 781 ACCGCTTTGAG 840  
DB 811 ACCGCTTTGAG 870  
QY 841 GCGCTTCTGAG 900  
DB 871 GCGCTTCTGAG 930  
QY 901 GAAAGAAATGCTGAG 960  
DB 931 GAAAGAAATGCTGAG 990  
QY 961 CTGCGCTTGAAGCTCTGAG 1020  
DB 991 CTGCGCTTGAAGCTCTGAG 1050  
QY 1021 GGGGAG 1080  
DB 1051 GGGGAG 1110  
QY 1081 CACACCCCTCTGAG 1140  
DB 1111 CACACCCCTCTGAG 1170  
QY 1141 GGGAG 1200  
DB 1171 GGGAG 1230  
QY 1201 AGTCTCGGAG 1260  
DB 1231 AGTCTCGGAG 1290  
QY 1261 GCGCTGAG 1320  
DB 1291 GCGCTGAG 1350

QY 1321 AGGACAGACAG 1380  
DB 1351 AGGACAGACAG 1410  
QY 1381 ACTGAG 1440  
DB 1411 ACTGAG 1470  
QY 1441 GAG 1500  
DB 1471 GAG 1530  
QY 1501 TTTTAAATTAATTTATATAT 1522  
DB 1531 TTTTAAATTAATTTATATAT 1552

## RESULT 11

ABX76124

ID ABX76124 standard; DNA; 1552 BP.

ABX76124;

02-APR-2003 (first entry)

Lung cancer-associated polynucleotide #1.

Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;  
antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.

Unidentified.

WO200286443-A2.

31-OCT-2002.

18-APR-2002; 2002MO-US012476.

18-APR-2001; 2001US-0284770P.

10-MAY-2001; 2001US-0290492P.

09-NOV-2001; 2001US-0339245P.

13-NOV-2001; 2001US-0350666P.

29-NOV-2001; 2001US-0334370P.

12-APR-2002; 2002US-0372246P.

(EOSB-) EOS BIOTECHNOLOGY INC.

Aziz N. Murray R;

WPI; 2003-093161/08.

P-PsDB; ABUS6408.

Detecting a lung cancer-associated transcript in a cell from a patient

for treating lung cancer, by contacting a biological sample from the

patient with a polynucleotide that exhibits increased or decreased

expression in lung cancer.

Claim 22; Page 189; 453bp; English.

The invention relates to a method for detecting a lung cancer-associated  
transcript in a cell from a patient, comprising contacting a biological  
sample from the patient with a polynucleotide that selectively hybridizes  
to a sequence that is at least 80 % identical to a gene that exhibits  
increased or decreased expression in lung cancer samples. Lung cancer-  
associated polynucleotides and polypeptides are used for identifying a  
compound that modulates a lung cancer-associated polypeptide, for  
inhibiting proliferation of a lung cancer-associated cell to treat lung  
cancer in a patient and for treating a mammal having lung cancer by  
administering a modulatory compound identified. The methods are useful  
for treating lung cancer, such as small cell lung cancer, non-small cell







XX MN CDNA clone.

XX MN; endogenous; Matu; quasi-viral agent; human; mammary tumour;

KW classical virus; slow virus; prion; exogenous MX; p58X;

KW cytoplasmic antigen; conservative; Hela cell; twin protein; p54/58N;

KW cell surface; nucleus; monoclonal antibody; Mab W75; neoplasm;

KW pre-neoplastic disease; vaccine; ss.

XX Homo sapiens.

OS

XX Key Location/Qualifiers

FT 1..1290

FT CDS /tag= a

XX MO9318152-A1.

PN 16-SEP-1993.

PD 08-MAR-1993; 93WO-US002024.

XX 11-MAR-1992; 92CS-00000709.

PR 21-OCT-1992; 92US-00964589.

XX (CIBA) CIBA CORNING DIAGNOSTICS CORP.

PA (VIRO-) INST VIROLOGY.

PI Zavada J, Pastorekova S, Pastorek J;

XX WPI: 1993-303466/38.

DR P-PSDB; AAR41746.

XX New MN gene and polypeptide(s) - used in diagnosis, prognosis and therapy

PT of neoplastic and/or pre-neoplastic disease.

XX Claim 2; Fig 1; 72pp; English.

XX This sequence represents the intronless MN gene which is a cellular gene

CC which is the endogenous component of the Matu agent. Matu is a novel

CC quasi-viral agent with rather unusual properties. It is presumably

CC derived from a human mammary tumour. In some aspects it resembles

CC classical viruses, whereas in other respects it resembles "slow" viruses

CC (prions), and in still other aspects it is different from both classes of

CC viruses. Matu is a two component system. One part of the complex, which

CC exogenous MX, is transmissible, and is manifest by a protein, p58X, which

CC is a cytoplasmic antigen which reacts with some natural sera, of humans

CC and of various animals. The other component, MN, is endogenous to human

CC cells. MN is a cellular gene showing very little homology with known DNA

CC sequences. It is rather conservative and present as a single copy in the

CC chromosomal DNA of various vertebrates. MN is manifest in Hela cells by a

CC nuclear protein p54/58N, that is localised on the cell surface and in the

CC nucleus. Immunoblot using a monoclonal antibody reactive with p54/58N

CC (Mab W75) reveals two bands at 54 kD and 58 kD. These two bands may

CC correspond to one type of protein that differs by glycosylation pattern

CC or by how it is processed. The expression of the MN gene is strongly

CC correlated with tumourigenicity. MN products can be used, in can be used

CC in diagnostic and/or prognostic assays for neoplastic and/or pre-

CC neoplastic disease. MN polypeptides, produced recombinantly by

CC unicellular hosts, can also be used for antibody production and in

CC vaccines for inducing protective immunity against neoplastic disease and

CC a dampening effect upon tumourigenic activity. (Updated on 25-MAR-2003 to

CC correct PN field.)

XX

SQ Sequence 1397 BP; 284 A; 402 C; 425 G; 286 T; 0 U; 0 Other;

Query Match 85.6%; Score 1302.8; DB 2; Length 1397;

Best Local Similarity 99.3%; Pred. No. 1.2e-307;

Matches 1392; Conservative 0; Mismatches 2; Indels 8; Gaps 8;

QY 124 CAGAGTTGCCCGGATGCAAGAGATTCCCTTGGAGAGAGGCTCTTCTGGGAGAT 183

Db 1 CAGAGTTGCCCGGATGCAAGAGATTCCCTTGGAGAGAGGCTCTTCTGGGAGAT 59

QY 184 GACCCACTGGCGAGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGATCCA 243

Db 60 GACCCACTGGCGAGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGATCCA 119

QY 244 CCCGAGAGAGAGATCTCACTGAGAGAGAGATCTCACTGAGAGAGAGATCTCACTGAA 303

Db 120 CCCGAGAGAGAGATCTCACTGAGAGAGAGATCTCACTGAGAGAGAGATCTCACTGAA 179

QY 304 GTTAA-GCCTAATCAGAGAGAGAGGCTCCCTGAAAGTTAGAGATCTCACTGTTGA 362

Db 180 GTTAAATGCTAATCAGAGAGAGAGGCTCCCTGAAAGTTAGAGATCTCACTGTTGA 239

QY 363 GGCCTCGAGATCTCAAGAACCCAGATTAATGCCCAGAGAGAGAGAGATGA 422

Db 240 GGCCTCGAGATCTCAAGAACCCAGATTAATGCCCAGAGAGAGAGAGATGA 299

QY 423 CCAGATCATTTGGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482

Db 300 CCAGATCATTTGGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 358

QY 483 GGGCCGCTTCCAGTCCCGGTGATATCCGCCCCAGCTGCGCGCTTCCGCGCGCT 542

Db 359 GGGCCGCTTCCAGTCCCGGTGATATCCGCCCCAGCTGCGCGCTTCCGCGCGCT 418

QY 543 GGGCCGCTTCCAGTCCCGGTGATATCCGCCCCAGCTGCGCGCTTCCGCGCGCT 602

Db 419 GGGCCGCTTCCAGTCCCGGTGATATCCGCCCCAGCTGCGCGCTTCCGCGCGCT 477

QY 603 CAATGGCCAGATGTCATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662

Db 478 CAATGGCCAGATGTCATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 537

QY 663 GCGGAGATACCGGGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722

Db 538 GCGGAGATACCGGGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 596

QY 723 GGAGCAACTGTGAG 782

Db 597 GGAGCAACTGTGAG 656

QY 783 GGCCTTTCAGAGTGCAG 842

Db 657 GGCCTTTCAGAGTGCAG 715

QY 843 CTTTGTGAG 901

Db 716 CTTTGTGAG 775

QY 902 AAGAAATGCTGAG 961

Db 776 AAGAAATGCTGAG 835

QY 962 TCCCTCTGACTTACGCGCTTCAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021

Db 836 TCCCTCTGACTTACGCGCTTCAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 895

QY 1022 CCCAGAGTCACTGAG 1081

Db 896 CCCAGAGTCACTGAG 955

QY 1082 AACACCTCTGAG 1141

Db 956 AACACCTCTGAG 1015

QY 1142 CGAGCAGAGCTTTGAATGAG 1201

Db 1016 CGAGCAGAGCTTTGAATGAG 1075

QY 1202 GTCCTCGAGCTCTGAG 1261

Db 1076 GTCCTCGAGCTCTGAG 1135

QY 1262 CCTGTGTTTGGCTCTTTTGTCTGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1321

Db 1136 CCTGTGTTTGGCTCTTTTGTGTGCAACAGCTCGGCTTCTGTGTCAATAGAA 1195  
Qy 1322 GGCACACAGAAAGGGAACCAAGGGGTGTGAC-TACGCCACAGAGGTAGCCGAG 1380  
Db 1156 GGCACACAGAAAGGGAACCAAGGGGTGTGACCGCCAGAGAGGTAGCCGAG 1255  
Qy 1381 ACTGAGGCTAGAGGCTGATCTTGAGAAATGTAGAGAGCAGAGCATCTGAGG 1440  
Db 1256 ACTGAGGCTAGAGGCTGATCTTGAGAAATGTAGAGAGCAGAGCATCTGAGG 1315  
Qy 1441 GAGCCGGTAACTGTCTCTGCTGCTATATGCACTCTTTAACTGCCAAGAAAT 1500  
Db 1316 GAGCCGGTAACTGTCTCTGCTGCTATATGCACTCTTTAACTGCCAAGAAAT 1375  
Qy 1501 TTTTAAATTAATTTATTAAT 1522  
Db 1376 TTTTAAATTAATTTATTAAT 1397

RESULT 14  
ID ADB53977 standard; DNA; 6521 BP.  
XX ADB53977;  
AC ADB53977;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE MNCA9 genomic DNA region.  
XX  
KM colon cell proliferative disorder; non methylated CpG dinucleotide;  
XX cytosinetic; cancer; adenoma; carcinoma; cytosine methylation state; ds.  
OS Unidentified.  
XX  
PN WO2003072821-A2.  
XX  
PD 04-SEP-2003.  
XX  
PF 27-FEB-2003; 2003WO-EP002035.  
XX  
PR 27-FEB-2002; 2002EP-00004551.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Adorjan P, Burger M, Maier S, Nimmich I, Becker E, Lesche R;  
PI Rujan T, Schmitt A;  
DR WPI; 2003-731620/69.  
XX  
PT Detecting and differentiating between colon cell proliferative disorders  
PT associated with a gene or its regulatory regions comprises contacting a  
PT target nucleic acid in a biological sample obtained from the subject with  
PT a reagent.  
XX  
PS Claim 46; SEQ ID NO 33; 74pp; English.  
XX  
CC The invention relates to a novel method for detecting and differentiating  
CC between colon cell proliferative disorders associated with at least one  
CC gene or its regulatory regions. The method comprises contacting a target  
CC nucleic acid in a biological sample obtained from the subject with at  
CC least one reagent or a series of reagents, where the reagent or series of  
CC reagents, distinguishes between methylated and non methylated CpG  
CC dinucleotides within the target nucleic acid. The molecules of the  
CC invention demonstrate cytosinetic activity whilst the method may useful  
CC for detecting and differentiating between colon cell proliferative  
CC disorders, including cancers such as colon adenoma and colon carcinoma.  
CC The PNA (peptide nucleic acid)-oligonucleotides are useful as probes for  
CC determining cytosine methylation state or single nucleotide  
CC polymorphisms. The current sequence is that of the genomic DNA region of  
CC taken from Wipoweb.  
XX

SEQ Sequence 6521 BP; 1650 A; 1570 C; 1641 G; 1660 T; 0 U; 0 Other;  
Query Match 27.2%; Score 414.4; DB 9; Length 6521;  
Best Local Similarity 99.8%; Pred. No. 1.2e-90;  
Matches 415; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ACAATCAGCCGATGAGCTCCCTGTGTCGCCACAGCCCTGAGCTCTGTGATCCGAGC 60  
Db 4931 ACAATCAGCCGATGAGCTCCCTGTGTCGCCACAGCCCTGAGCTCTGTGATCCGAGC 4990  
Qy 61 CTTGCTCAGGCTCTCACTGTGCAACTGCTGCTGTCACTGCTCTTGTATGCTTCCAT 120  
Db 4991 CTTGCTCAGGCTCTCACTGTGCAACTGCTGCTGTCACTGCTCTTGTATGCTTCCAT 5050  
Qy 121 CCCAGAGGTTGCCCGGATGAGAGATTTCCCTTGGAGAGAGGCTCTTGGAGAA 180  
Db 5051 CCCAGAGGTTGCCCGGATGAGAGATTTCCCTTGGAGAGAGGCTCTTGGAGAA 5110  
Qy 181 GATGACCCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCAACCAAGAGAGAGAT 240  
Db 5111 GATGACCCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCAACCAAGAGAGAGAT 5170  
Qy 241 CCACCCGAGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 300  
Db 5171 CCACCCGAGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 5230  
Qy 301 GAAGTTAAGCTTAATCAGAGAGAGAGGCTCCCTGAAGTTAGAGATCTACTGAT 360  
Db 5231 GAAGTTAAGCTTAATCAGAGAGAGAGGCTCCCTGAAGTTAGAGATCTACTGAT 5290  
Qy 361 GAGGCTCTGAGATCTCTCAAGAACCCGAGATTAATGCCACAGAGAGAGAGAGAG 416  
Db 5291 GAGGCTCTGAGATCTCTCAAGAACCCGAGATTAATGCCACAGAGAGAGAGAGAG 5346

RESULT 15  
ID AAT09187 standard; DNA; 10897 BP.  
XX AAT09187  
AC AAT09187;  
XX  
DT 25-MAR-2003 (revised)  
XX  
DE 26-JUL-1996 (first entry)  
XX  
DE Mutu putative oncogene MN genomic DNA.  
XX  
KM Mutu; endogenous; cellular component; MN; HeLa cell; diagnosis;  
XX Lymphocytic choriomeningitis virus; LCMV; putative oncogene; treatment;  
XX neoplastic; pre-neoplastic; disease; antisense therapy; antibody;  
XX vaccine; vertebrate; immunisation; ss.  
OS Homo sapiens.  
XX  
FH Key  
FH misc\_feature  
FT 1973  
FT Location/Qualifiers  
FT /tag= a  
FT /note= "Given as "P" in specification"  
FT 3536  
FT /tag= b  
FT /number= 1  
FT 5125..5154  
FT /tag= c  
FT /number= 2  
FT 5350..5518  
FT /tag= d  
FT /number= 3  
FT 5650..5792  
FT /tag= e  
FT /number= 4  
FT 5882..5974  
FT /tag= f  
FT /number= 5  
FT 7375..7441  
FT exon





OY	1	AAGTCAGCCGAGATGGGCTCCGCTGTGGCCCAAGGCCCTGGCTCCCTGTGTATCCCGAGC	60
Db	1	ACAGTCAGCCGAGATGGGCTCCGCTGTGGCCCAAGGCCCTGGCTCCCTGTGTATCCCGAGC	60
OY	61	CCTGCTCCAGGCGCTCACTGTGCCAATGCTGCTGTCACTGCTGCTTCTGTATGCTGTCCAT	120
Db	61	CCTGCTCCAGGCGCTCACTGTGCCAATGCTGCTGTCACTGCTGCTTCTGTATGCTGTCCAT	120
OY	121	CCCTCAGAGATTCCCCGGATGCAAGAGATTCCCCCTGTGGAGAGAGGCTCTTCTGGGAAA	180
Db	121	CCCTCAGAGATTCCCCGGATGCAAGAGATTCCCCCTGTGGAGAGAGGCTCTTCTGGGAAA	180
OY	181	GATGACCCACTGGGCGAGAGAGATCTGCTCCAGTGAAAGAGATTCAACCAAGAGAGAT	240
Db	181	GATGACCCACTGGGCGAGAGAGATCTGCTCCAGTGAAAGAGATTCAACCAAGAGAGAT	240
OY	241	CCATCCCGAGAGAGAGATCTACTCTGGAAGAGATCTACTGAGAGAGAGATCTACT	300
Db	241	CCATCCCGAGAGAGAGATCTACTCTGGAAGAGATCTACTGAGAGAGAGATCTACT	300
OY	301	GAAGTTAAGCTTAAATCAAGAAAGAGGCTCCCTGAATTAAGAGATCTACTACTGT	360
Db	301	GAAGTTAAGCTTAAATCAAGAAAGAGGCTCCCTGAATTAAGAGATCTACTACTGT	360
OY	361	GAGGCTCCCTGAGATCTCTCAAGAACCCCGAATATATCCCAACAGGATCAAGAGAGGAT	420
Db	361	GAGGCTCCCTGAGATCTCTCAAGAACCCCGAATATATATCCCAACAGGATCAAGAGAGGAT	420
OY	421	GACCAAGTCAATTGGCGCTATGAGAGGAGCCCGGCTTGACCTCCGGGTGTCCCAAGCTTGC	480
Db	421	GACCAAGTCAATTGGCGCTATGAGAGGAGCCCGGCTTGACCTCCGGGTGTCCCAAGCTTGC	480
OY	481	GCGAGCGGCTTCCAGTCCCGGGTGAATATCCGCCCCAGAGCTGCGCGCTTCTGCGCGGAC	540
Db	481	GCGAGCGGCTTCCAGTCCCGGGTGAATATCCGCCCCAGAGCTGCGCGCTTCTGCGCGGAC	540
OY	541	CTGCGCCCTTGGAACTCTGTGGGCTTCCAGCTCCCGCTCCCAAGCTGCGCTTGGC	600
Db	541	CTGCGCCCTTGGAACTCTGTGGGCTTCCAGCTCCCGCTCCCAAGCTGCGCTTGGC	600
OY	601	AACATATGGGCAACAGTGTGAATGAATGACCTGTGCTGTGGCTAAGATATGCTCTGGTCC	660
Db	601	AACATATGGGCAACAGTGTGAATGAATGACCTGTGCTGTGGCTAAGATATGCTCTGGTCC	660
OY	661	GCGCGGAGTACCGGCGCTTGTCACTGTGAATGTGAATCTGGAGCTGGAGGCTGTGCTGGGAC	720
Db	661	GCGCGGAGTACCGGCGCTTGTCACTGTGAATGTGAATCTGGAGCTGGAGGCTGTGCTGGGAC	720
OY	721	TCCGAGACACATGTGAGAGGACACCGTTTCCCTCCAGATTCACAGTGTTCACCTCAGC	780
Db	721	TCCGAGACACATGTGAGAGGACACCGTTTCCCTCCAGATTCACAGTGTTCACCTCAGC	780
OY	781	ACGCGCTTTGCCAGATTTACAGAGGCTTGGGGGCGCCCGGAGAGGCTTGGCGATTTGGCC	840
Db	781	ACGCGCTTTGCCAGATTTACAGAGGCTTGGGGGCGCCCGGAGAGGCTTGGCGATTTGGCC	840
OY	841	GCTTTTGTGAGAGAGGCGCCGGAATAAACAAGTCCATATAGCAAGTTGTCTGTGCTTG	900
Db	841	GCTTTTGTGAGAGAGGCGCCGGAATAAACAAGTCCATATAGCAAGTTGTCTGTGCTTG	900
OY	901	GAGAAATGCTTGAAGAGAGGCTCAAGACTCAAGTCCCAAGACTGGAACATATCTGACATC	960
Db	901	GAGAAATGCTTGAAGAGAGGCTCAAGACTCAAGTCCCAAGACTGGAACATATCTGACATC	960
OY	961	CTGCGCTTCACTTACGCGCTACTTCCAAATATAGGGGGTCTCTGACTACACCGCGCTGT	1020
Db	961	CTGCGCTTCACTTACGCGCTACTTCCAAATATAGGGGGTCTCTGACTACACCGCGCTGT	1020
OY	1021	GCCCAAGGATCTCTGAGATGTTTAAACAAGACAGTATGCTGATGCTGAAGAGCTC	1080
Db	1021	GCCCAAGGATCTCTGAGATGTTTAAACAAGACAGTATGCTGATGCTGAAGAGCTC	1080

Qy	1081	CACACCCCTCTGACACCCCTGCTGGGGACCGGGGACTCTCGAGCTGACCTGAACCTCCGA	1140
Db	1081	CACACCCCTCTGACACCCCTGCTGGGGACCTGGTGACTCTGAGCTACAGCTGAACCTCCGA	1140
Qy	1141	GGAGCGACACCTTTGAAATGGGCGAGTGAATGAGAGCCTCTTTCCTGCTGAGTGAACAC	1200
Db	1141	GGAGCGACACCTTTGAAATGGGCGAGTGAATGAGAGCCTCTTTCCTGCTGAGTGAACAGC	1200
Qy	1201	AGTCTCTGGGCTGCTGAGCCAGTCCAGCTGAAATTCCTGCTGGGCTGAGTGAACATCCTTA	1260
Db	1201	AGTCTCTGGGCTGCTGAGCCAGTCCAGCTGAAATTCCTGCTGGGCTGAGTGAACATCCTTA	1260
Qy	1261	GCCTCGATTTTGGGCTCTTTTGCTGCTGACACAGCGTGCCTTCTTGTGAGATGAGA	1320
Db	1261	GCCTCGATTTTGGGCTCTTTTGCTGCTGACACAGCGTGCCTTCTTGTGAGATGAGA	1320
Qy	1321	AGGCAACACAGAGGGGAAACCAAGGGGGTGTAGTACTACCGCCACAGCAGATGACGAG	1380
Db	1321	AGGCAACACAGAGGGGAAACCAAGGGGGTGTAGTACTACCGCCACAGCAGATGACCGAG	1380
Qy	1381	ACTGAGACCTTAGAGCTGATCTTGAGAAATGAGAAACCCAGCCAGAGGACATCTGAGGG	1440
Db	1381	ACTGAGACCTTAGAGCTGATCTTGAGAAATGAGAAACCCAGCCAGAGGACATCTGAGGG	1440
Qy	1441	GGAGCGGTAACGTGCTCTGCTGCTCATTTATGCCACTTCTTTTAATCTGCCAAGAAATT	1500
Db	1441	GGAGCGGTAACGTGCTCTGCTCATTTATGCCACTTCTTTTAATCTGCCAAGAAATT	1500
Qy	1501	TTTTAAATTAATATTTTAAT	1522
Db	1501	TTTTAAATTAATATTTTAAT	1522

RESULT 2  
US-C8-477-504A-1

Sequence 1, Application US/08477504A  
; Patent No. 5972353

GENERAL INFORMATION:  
APPLICANT: Zavada, Jan

APPLICANT: Závada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir

APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and  
NUMBER OF SEQUENCES: 85

NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court

CITY: Tiburon  
STATE: California

STATE: California  
COUNTRY: USA  
ZIP: 94920

```

; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDICAL TYPE FLOppy disk

```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: DOS/MS-DOS

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/477,

```

APPLICATION NUMBER: 05/06/4111  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260

APPLICATION NUMBER: US 08/260,  
FILING DATE: 15-JUN-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.

REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-002

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034

TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEO ID NO.: 1.

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;

```

LENGTH: 1522 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-477-504A-1

Query Match 100.0%; Score 1522; DB 2; Length 1522;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTCAGCCGATGAGCTCCCTGTCGCCCCAGCCCCCTGCTCCCTGTTGATCCCGGCC 60  
 Db 1 ACAGTCAGCCGATGAGCTCCCTGTCGCCCCAGCCCCCTGCTCCCTGTTGATCCCGGCC 60  
 QY 61 CCTGCTCCAGGCTCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 Db 61 CCTGCTCCAGGCTCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 QY 121 CCCAGAGGTTGCCCCGAGTGCAGAGAGATTCCCTTGAGAGAGGCTCTTCTGAGGAA 180  
 Db 121 CCCAGAGGTTGCCCCGAGTGCAGAGAGATTCCCTTGAGAGAGGCTCTTCTGAGGAA 180  
 QY 181 GATGACCCATGCGGCGAGAGAGATCTGCCAGTGAAGAGATTCAACCAGAGAGAGAGAT 240  
 Db 181 GATGACCCATGCGGCGAGAGAGATCTGCCAGTGAAGAGATTCAACCAGAGAGAGAGAT 240  
 QY 241 CCACCCGAGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACCT 300  
 Db 241 CCACCCGAGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACCT 300  
 QY 301 GAGTTAAGCCTTAATCAGAGAAGAGGCTCCCTGAAATTGAGAGATCTACCTGCTT 360  
 Db 301 GAGTTAAGCCTTAATCAGAGAAGAGGCTCCCTGAAATTGAGAGATCTACCTGCTT 360  
 QY 361 GAGGCTCTGAGAGATCTCAAGAACCCAGAAATTAAGCCAGAGAGAGAGAGAGAGAT 420  
 Db 361 GAGGCTCTGAGAGATCTCAAGAACCCAGAAATTAAGCCAGAGAGAGAGAGAGAGAT 420  
 QY 421 GACAGAGATATTGGCGCTAGTGAAGCGACCCGCTGAGCCCGGCTGCCAGACTGC 480  
 Db 421 GACAGAGATATTGGCGCTAGTGAAGCGACCCGCTGAGCCCGGCTGCCAGACTGC 480  
 QY 481 GCGGCGCGCTTCCAGTCCCGGCTGATATCCGCCCCAGCTGCGCGCTTTCGCCCGCC 540  
 Db 481 GCGGCGCGCTTCCAGTCCCGGCTGATATCCGCCCCAGCTGCGCGCTTTCGCCCGCC 540  
 QY 541 CTGCGCCCTGGAACTCTGGGCTTCCAGCTCCCGCGCTGCCAGAACTGCGCTGCGC 600  
 Db 541 CTGCGCCCTGGAACTCTGGGCTTCCAGCTCCCGCGCTGCCAGAACTGCGCTGCGC 600  
 QY 601 AACATAGGCGACAGTGTGCACTGACCTGCTCTGAGCTAGATGGCTCTGAGTCC 660  
 Db 601 AACATAGGCGACAGTGTGCACTGACCTGCTCTGAGCTAGATGGCTCTGAGTCC 660  
 QY 661 GGGCGGAGATACCGGGCTCTGAGCTGCACTGCACTGGGGGCTGCAAGTCTGCGGGC 720  
 Db 661 GGGCGGAGATACCGGGCTCTGAGCTGCACTGCACTGGGGGCTGCAAGTCTGCGGGC 720  
 QY 721 TCGGAGCACTGTGAGAGGCAACGCTTCCCTGCGAGATTCACGTGCTTCACTCAAC 780  
 Db 721 TCGGAGCACTGTGAGAGGCAACGCTTCCCTGCGAGATTCACGTGCTTCACTCAAC 780  
 QY 781 ACCGCTTTGCCAGATTGACGAGGCTTGGGGGCGCCGAGAGGCTGAGCGTGTGAGCC 840  
 Db 781 ACCGCTTTGCCAGATTGACGAGGCTTGGGGGCGCCGAGAGGCTGAGCGTGTGAGCC 840  
 QY 841 GCCTTTCTGAGAGAGGCGCGGAGAAACAGTGCCTTATGAGCACTTCTCTGCTG 900  
 Db 841 GCCTTTCTGAGAGAGGCGCGGAGAAACAGTGCCTTATGAGCACTTCTCTGCTG 900  
 QY 901 GAAGAATGCTAGAGAGGCTCAGAGATCCAGAGCTGAGACATATCTGCACTC 960

Db 901 GAAGAATGCTAGAGAGGCTCAGAGATCCAGAGCTGAGTCCAGAGCTGACATATCTGCACTC 960  
 QY 961 CTGCCCCCTGACTTACAGCCGCTACTTCCAAATATAGAGGCTCTGATATACCGCCCTGT 1020  
 Db 961 CTGCCCCCTGACTTACAGCCGCTACTTCCAAATATAGAGGCTCTGATATACCGCCCTGT 1020  
 QY 1021 GCCCAGGCTGATCTGAGTGTGTTTAAACAGACAGTGAATGCTGAGTGAAGCACTC 1080  
 Db 1021 GCCCAGGCTGATCTGAGTGTGTTTAAACAGACAGTGAATGCTGAGTGAAGCACTC 1080  
 QY 1081 CACACCTCTCTGACACCTCTGTGAGACCTGTGATCTCTGCTACAGCTGAACCTTCCGA 1140  
 Db 1081 CACACCTCTCTGACACCTCTGTGAGACCTGTGATCTCTGCTACAGCTGAACCTTCCGA 1140  
 QY 1141 GCGAGCAGCCTTTGAATGAGGCGAGTATGAGGAGCTCTTCCCTGCTGAGTGAAGCAAGC 1200  
 Db 1141 GCGAGCAGCCTTTGAATGAGGCGAGTATGAGGAGCTCTTCCCTGCTGAGTGAAGCAAGC 1200  
 QY 1201 AGTCTCGGCTGCTGAGCCAGTCCAGCTGAATTCCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 Db 1201 AGTCTCGGCTGCTGAGCCAGTCCAGCTGAATTCCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 QY 1261 GCGCTGCTTTTGGCTCCTTTTGTGCTGACACGAGCTCGGCTTCTGCTGCTGCTGCTG 1320  
 Db 1261 GCGCTGCTTTTGGCTCCTTTTGTGCTGACACGAGCTCGGCTTCTGCTGCTGCTGCTG 1320  
 QY 1321 AGGCAACAAGAGAGGAGAACCAAGAGGAGTGTGAGCTACCCGCGAGAGAGTACCGGAG 1380  
 Db 1321 AGGCAACAAGAGAGGAGAACCAAGAGGAGTGTGAGCTACCCGCGAGAGAGTACCGGAG 1380  
 QY 1381 ACTGAGGCTGAGGCTGAGTCTGAGAAATGAGAAAGCAGCAGAGCATCTGAGG 1440  
 Db 1381 ACTGAGGCTGAGGCTGAGTCTGAGAAATGAGAAAGCAGCAGAGCATCTGAGG 1440  
 QY 1441 GAGGCGGTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
 Db 1441 GAGGCGGTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
 QY 1501 TTTTAAATTAATTTTAAAT 1522  
 Db 1501 TTTTAAATTAATTTTAAAT 1522

RESULT 3  
 US-08-486-756A-1  
 Sequence 1, Application US/08486756A  
 Patent No. 5981711  
 GENERAL INFORMATION:  
 APPLICANT: Zavada, Jan  
 APPLICANT: Pastorekova, Silvia  
 APPLICANT: Pastorek, Jaromir  
 TITLE OF INVENTION: MN Gene and Protein  
 NUMBER OF SEQUENCES: 86  
 CORRESPONDENCE ADDRESS:  
 ADDRESSES: Leona L. Lauder  
 STREET: 6 Mariposa Court  
 CITY: Tiburon  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94920  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/486,756A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION NUMBER: US 08/260,190  
 FILING DATE: 15-JUN-1994  
 ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.  
 REGISTRATION NUMBER: 30,863  
 REFERENCE/DOCKET NUMBER: D-0021.3C  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-435-2034  
 TELEFAX: 415-435-0727  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1522 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-486-756A-1

Query Match 100.0%; Score 1522; DB 2; Length 1522;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGATGAGCCGATGAGTCCCTGTCGCCAGCCCGTGGCTCCCTGTGTGATCCGAGCC 60  
 DB 1 AAGATGAGCCGATGAGTCCCTGTCGCCAGCCCGTGGCTCCCTGTGTGATCCGAGCC 60  
 QY 61 CCGTCTCAGAGCCTCACTGTGCACTGTGCTGTCTCACTGTGCTGTGTGATGCTGTGAT 120  
 DB 61 CCGTCTCAGAGCCTCACTGTGCACTGTGCTGTCTCACTGTGCTGTGTGATGCTGTGAT 120  
 QY 121 CCCGAGAGTGGCCCGGATGAGAGGATTTCCCTTGGAGAGAGGCTCTTCTGAGGAA 180  
 DB 121 CCCGAGAGTGGCCCGGATGAGAGGATTTCCCTTGGAGAGAGGCTCTTCTGAGGAA 180  
 QY 181 GATGACCCACTGAGGAGAGAGATCTGCCCAGTGAAGAGATTACCCAGAGAGAGAT 240  
 DB 181 GATGACCCACTGAGGAGAGAGATCTGCCCAGTGAAGAGATTACCCAGAGAGAGAT 240  
 QY 241 CCACCCGAGAGAGAGATCTACTGTGAGAGAGATCTTACTGTGAGAGAGATCTTACT 300  
 DB 241 CCACCCGAGAGAGAGATCTACTGTGAGAGAGATCTTACTGTGAGAGAGATCTTACT 300  
 QY 301 GAAGTTAAGCCTTAATCAGAGAGAGAGGCTCCCTGAAGTTAGAGATCTTACTGTGTT 360  
 DB 301 GAAGTTAAGCCTTAATCAGAGAGAGAGGCTCCCTGAAGTTAGAGATCTTACTGTGTT 360  
 QY 361 GAGGCTCTTGAAGATCTCTAAGAACCCAGATTAATGCCCAAGAGAGAGAGAGAT 420  
 DB 361 GAGGCTCTTGAAGATCTCTAAGAACCCAGATTAATGCCCAAGAGAGAGAGAT 420  
 QY 421 GACAGAGATCATTTGAGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480  
 DB 421 GACAGAGATCATTTGAGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480  
 QY 481 GCGGCGCGCTTCCAGTCCCGGATGATATCCGCCCCAGCTGCGCCCTTCTGCGCGCC 540  
 DB 481 GCGGCGCGCTTCCAGTCCCGGATGATATCCGCCCCAGCTGCGCCCTTCTGCGCGCC 540  
 QY 541 CTGCGCCCGCTGGAATCTCTGAGGCTTCCAGCTCCCGCGCTCCAGAACTGCGCTCGC 600  
 DB 541 CTGCGCCCGCTGGAATCTCTGAGGCTTCCAGCTCCCGCGCTCCAGAACTGCGCTCGC 600  
 QY 601 AACATGAGCCACAGTGTGCACTGACCTGCTCTGAGGCTGAGAGATGCTGTGCTGCC 660  
 DB 601 AACATGAGCCACAGTGTGCACTGACCTGCTCTGAGGCTGAGAGATGCTGTGCTGCC 660  
 QY 661 GGGGAGAGATACCGGGCTCTGAGAGCTGCACTGCACTGAGAGAGAGAGAGAGAGAT 720  
 DB 661 GGGGAGAGATACCGGGCTCTGAGAGCTGCACTGCACTGAGAGAGAGAGAGAGAT 720  
 QY 721 TCGAGACACACTGTGAT 780  
 DB 721 TCGAGACACACTGTGAT 780

QY 781 ACCGCTTTGACAGAGTTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 840  
 DB 781 ACCGCTTTGACAGAGTTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 840  
 QY 841 GCGTTTGTGAT 900  
 DB 841 GCGTTTGTGAT 900  
 QY 901 GAAGAATTCCTGAT 960  
 DB 901 GAAGAATTCCTGAT 960  
 QY 961 CTGCGCTCTGACTTCAAGCGGCTACTTCCAAATATGAGAGAGAGAGAGAGAGAGAT 1020  
 DB 961 CTGCGCTCTGACTTCAAGCGGCTACTTCCAAATATGAGAGAGAGAGAGAGAGAGAT 1020  
 QY 1021 GCGGAGAGATCATCTGAGATCTGAGATCTGAGATCTGAGATCTGAGATCTGAGATCT 1080  
 DB 1021 GCGGAGAGATCATCTGAGATCTGAGATCTGAGATCTGAGATCTGAGATCTGAGATCT 1080  
 QY 1081 CACACCTCTCTGACACCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140  
 DB 1081 CACACCTCTCTGACACCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140  
 QY 1141 GCGAGAGAGCTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1200  
 DB 1141 GCGAGAGAGCTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1200  
 QY 1201 AGTCTGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1260  
 DB 1201 AGTCTGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1260  
 QY 1261 GCGCTGTTTTTGGCTCTCTTTTGTCTGTCAACAGAGAGAGAGAGAGAGAGAT 1320  
 DB 1261 GCGCTGTTTTTGGCTCTCTTTTGTCTGTCAACAGAGAGAGAGAGAGAGAGAT 1320  
 QY 1321 AGCGAT 1380  
 DB 1321 AGCGAT 1380  
 QY 1381 ACTGAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1440  
 DB 1381 ACTGAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1440  
 QY 1441 GAGGCGGTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
 DB 1441 GAGGCGGTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
 QY 1501 TTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1522  
 DB 1501 TTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1522

RESULT 4  
 US-08-485-862B-1  
 ; Sequence 1, Application US/08485862B  
 ; Patent No. 5989838  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zavada, Jan  
 ; APPLICANT: Pastorekova, Silvia  
 ; APPLICANT: Pastorek, Jaromir  
 ; TITLE OF INVENTION: MY Gene and Protein  
 ; NUMBER OF SEQUENCES: 86  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Leona L. Lauder  
 ; STREET: 5 Mariposa Court  
 ; CITY: Tiburon  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94920  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/485,862B  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/477,504  
 FILING DATE: 07-JUN-1995  
 APPLICATION NUMBER: US 08/260,190  
 FILING DATE: 15-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lauder, Leona L.  
 REGISTRATION NUMBER: 30,863  
 REFERENCE/DOCKET NUMBER: D-0021.3D  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-435-2034  
 TELEFAX: 415-435-0727  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1522 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-485-862B-1

Query Match 100.0%; Score 1522; DB 2; Length 1522;  
 Beet Local Similarity 100.0%; Pred. No. 0;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTACGCGGAGTGGTCTCCCTGTCGCCAGCCCTGCTCCCTGTTGATCCCGGCC 60  
 DB 1 ACAGTACGCGGAGTGGTCTCCCTGTCGCCAGCCCTGCTCCCTGTTGATCCCGGCC 60  
 QY 61 CCTGCTCCAGGCGCTCACTGTCGACACTGCTGTCACTGCTGTTGATGCTGCTGCT 120  
 DB 61 CCTGCTCCAGGCGCTCACTGTCGACACTGCTGTCACTGCTGTTGATGCTGCTGCT 120  
 QY 121 CCCGAGAGTGGTCCCGGATGCAAGAGATTCCTGAGAGAGAGGCTTTCTGAGGAA 180  
 DB 121 CCCGAGAGTGGTCCCGGATGCAAGAGATTCCTGAGAGAGAGGCTTTCTGAGGAA 180  
 QY 121 CCCGAGAGTGGTCCCGGATGCAAGAGATTCCTGAGAGAGAGGCTTTCTGAGGAA 180  
 DB 121 CCCGAGAGTGGTCCCGGATGCAAGAGATTCCTGAGAGAGAGGCTTTCTGAGGAA 180  
 QY 181 GATGACCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTACCCAGAGAGAGAT 240  
 DB 181 GATGACCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTACCCAGAGAGAGAT 240  
 QY 241 CCACCCGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 300  
 DB 241 CCACCCGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 300  
 QY 301 GAGTTAAGCTTAATCAAGAAAGAGGCTCCCTGAAGTTAAGATCTACTACTGTT 360  
 DB 301 GAGTTAAGCTTAATCAAGAAAGAGGCTCCCTGAAGTTAAGATCTACTACTGTT 360  
 QY 361 GAGGCTCTGAGATCTCAAGAAAGAGGCTCCCTGAAGTTAAGATCTACTACTGTT 420  
 DB 361 GAGGCTCTGAGATCTCAAGAAAGAGGCTCCCTGAAGTTAAGATCTACTACTGTT 420  
 QY 421 GACGAGATCATTTGGGCTATGAGAGGAGCCGCTGAGGCTCCGAGGCTGTC 480  
 DB 421 GACGAGATCATTTGGGCTATGAGAGGAGCCGCTGAGGCTCCGAGGCTGTC 480  
 QY 481 GAGGAGGCTTCAGTCCGAGTGAATTCGAGCCGAGCTGAGGCTTCCTGCGGAGC 540  
 DB 481 GAGGAGGCTTCAGTCCGAGTGAATTCGAGCCGAGCTGAGGCTTCCTGCGGAGC 540  
 QY 541 CTGCGGCTCTGAGAACTCTGAGGCTTCAGGCTCCGAGCTCCGAGAACTGCGCTGCGG 600  
 DB 541 CTGCGGCTCTGAGAACTCTGAGGCTTCAGGCTCCGAGCTCCGAGAACTGCGCTGCGG 600

QY 601 AACATGCGCAAGTGTGCACTGACCTGCTCCCTGAGCTAAGAGATGCTCTGAGTCC 660  
 DB 601 AACATGCGCAAGTGTGCACTGACCTGCTCCCTGAGCTAAGAGATGCTCTGAGTCC 660  
 QY 661 GGGCGGAGTACCGGGCTCTGCAAGTGTGCACTGAGGAGGCTGCAAGGCTCCGGGC 720  
 DB 661 GGGCGGAGTACCGGGCTCTGCAAGTGTGCACTGAGGAGGCTGCAAGGCTCCGGGC 720  
 QY 721 TCGAGACACATGTGAGAGGAGCCAGCTTCCCTGCGAGATCCAGTGTTCACCTCAGC 780  
 DB 721 TCGAGACACATGTGAGAGGAGCCAGCTTCCCTGCGAGATCCAGTGTTCACCTCAGC 780  
 QY 781 ACCGCTTTGCGAGTGTACAGAGGCTTGGGGGCGCCGGAGAGGCTGCGCTGTGCGC 840  
 DB 781 ACCGCTTTGCGAGTGTACAGAGGCTTGGGGGCGCCGGAGAGGCTGCGCTGTGCGC 840  
 QY 841 GCCTTCTGAGAGAGGCGCCGAGAGAAACATGCTCTATGACAGTGTCTGTGCTG 900  
 DB 841 GCCTTCTGAGAGAGGCGCCGAGAGAAACATGCTCTATGACAGTGTCTGTGCTG 900  
 QY 901 GAAGAAATGCTGAGAGAGGCTTCAAGTGTCCAGAGCTGAGCATATCTGCACTC 960  
 DB 901 GAAGAAATGCTGAGAGAGGCTTCAAGTGTCCAGAGCTGAGCATATCTGCACTC 960  
 QY 961 CTGCGCTCTGACTTCAAGGCTTCACTTCAATGAGGAGGCTCTGACTACACCGCTGT 1020  
 DB 961 CTGCGCTCTGACTTCAAGGCTTCACTTCAATGAGGAGGCTCTGACTACACCGCTGT 1020  
 QY 1021 GCCCAGGCTGTCACTGAGACTGTGTTAAACAGAGATGATCTGAGTCTTAAGAGCTC 1080  
 DB 1021 GCCCAGGCTGTCACTGAGACTGTGTTAAACAGAGATGATCTGAGTCTTAAGAGCTC 1080  
 QY 1081 CACACCTCTGTCACACCTGTCGAGGAGCTGTCGTCGCTGCTGAGTGAACCTCGGA 1140  
 DB 1081 CACACCTCTGTCACACCTGTCGAGGAGCTGTCGTCGCTGCTGAGTGAACCTCGGA 1140  
 QY 1141 GCGACGAGCTTTGATGAGGAGGAGTGAATGAGGCTCTTCCCTGCTGAGTGAACG 1200  
 DB 1141 GCGACGAGCTTTGATGAGGAGGAGTGAATGAGGCTCTTCCCTGCTGAGTGAACG 1200  
 QY 1201 AGTCTCTGAGGCTCTGAGGAGTGAATGAGGCTCTTCCCTGCTGAGTGAACG 1260  
 DB 1201 AGTCTCTGAGGCTCTGAGGAGTGAATGAGGCTCTTCCCTGCTGAGTGAACG 1260  
 QY 1261 GCGCTGTTTGGCTCTCTTTTGTGCTCACACAGGCTGCGCTTCTGTCAGATGAGA 1320  
 DB 1261 GCGCTGTTTGGCTCTCTTTTGTGCTCACACAGGCTGCGCTTCTGTCAGATGAGA 1320  
 QY 1321 AGGACGACAGAGGAGGAGCAAGAGGAGTGTGAGCTACCGCCGAGAGAGTACCGAG 1380  
 DB 1321 AGGACGACAGAGGAGGAGCAAGAGGAGTGTGAGCTACCGCCGAGAGAGTACCGAG 1380  
 QY 1381 ACTGAGAGCTTGAAGGCTGAGTCTTGAAGATGAGAGGAGGAGGAGTCTGAGAGG 1440  
 DB 1381 ACTGAGAGCTTGAAGGCTGAGTCTTGAAGATGAGAGGAGGAGGAGTCTGAGAGG 1440  
 QY 1441 GAGGCGGTAAGTGTCTGCTGCTGCTCATTAATGCACTTCTTTTAACTGCGCAAGAA 1500  
 DB 1441 GAGGCGGTAAGTGTCTGCTGCTGCTCATTAATGCACTTCTTTTAACTGCGCAAGAA 1500  
 QY 1501 TTTTAAATTAATTAATTAAT 1522  
 DB 1501 TTTTAAATTAATTAATTAAT 1522

RESULT 5  
 US-08-787-739-1  
 Sequence 1, Application US/0878739  
 Patent No. 6027887  
 GENERAL INFORMATION:  
 APPLICANT: Zavada, Jan  
 APPLICANT: Pastorekova, Silvia  
 APPLICANT: Pastorek, Jaromir







```

Db      961  CTGCGCTGACTTACGCCGCTACTTCCATATGAGGGTCTCTGACTACACCGCCCTGT 1020
QY      1021  GCCCAGGGTGTACTGTGACTGTGTTTAAACAGACAGTAGTCTGAGTGTCTAAGCAGCTC 1080
Db      1021  GCCCAGGGTGTACTGTGACTGTGTTTAAACAGACAGTAGTCTGAGTGTCTAAGCAGCTC 1080
QY      1081  CACACCTCTCTGACACCTGTGAGGAGCTGTGACTCTCGGCTCACTCACTAATCTCCGA 1140
Db      1081  CACACCTCTCTGACACCTGTGAGGAGCTGTGACTCTCGGCTCACTCACTAATCTCCGA 1140
QY      1141  GCGACGACGCTTTGATGAGGAGCTGATTTGAGGCTCTCTCCCTGCTGAGTGAACAGC 1200
Db      1141  GCGACGACGCTTTGATGAGGAGCTGATTTGAGGCTCTCTCCCTGCTGAGTGAACAGC 1200
QY      1201  AGTCTCTGGGCTCTGAGCAGTCCAGTGAATTTCTGCTGAGTCTGCTGAGTGAATCTTA 1260
Db      1201  AGTCTCTGGGCTCTGAGCAGTCCAGTGAATTTCTGCTGAGTCTGCTGAGTGAATCTTA 1260
QY      1261  GCCCTGATTTTGGCTCTCTTTTGTGCTGACACAGCGTCCGCTTCTGCTGAGTGAATGA 1320
Db      1261  GCCCTGATTTTGGCTCTCTTTTGTGCTGACACAGCGTCCGCTTCTGCTGAGTGAATGA 1320
QY      1321  AGCAGACACAGAGGGGAAACCAAGGGGTGTGAGCTACCGCCACAGAGTGTGCGAG 1380
Db      1321  AGCAGACACAGAGGGGAAACCAAGGGGTGTGAGCTACCGCCACAGAGTGTGCGAG 1380
QY      1381  ACTGAGACCTAGAGGCTGTGATCTTGAGAAATGTGAAAGCCAGCAGAGGATCTGAGG 1440
Db      1381  ACTGAGACCTAGAGGCTGTGATCTTGAGAAATGTGAAAGCCAGCAGAGGATCTGAGG 1440
QY      1441  GAGCGCGTACTGTCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Db      1441  GAGCGCGTACTGTCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY      1501  TTTTAAATTAATTTTAAAT 1522
Db      1501  TTTTAAATTAATTTTAAAT 1522

```

RESULT 8  
US-08-485-049D-1

```

; Sequence 1, Application US/08485049D
; Patent No. 6204370
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 369 Pine Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,049D
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3E
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 415-981-2034
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1522 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-485-049D-1

```

Query Match 100.0%; Score 1522; DB 3; Length 1522;

Best Local Similarity 100.0%; Pred. NO. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1  ACAGTCAAGCCGATGCTCCCTGTGCCCCAGCCCTGAGTCCCTCTGTTGATCCGAGC 60
Db      1  ACAGTCAAGCCGATGCTCCCTGTGCCCCAGCCCTGAGTCCCTCTGTTGATCCGAGC 60
QY      61  CTTGCTCAGAGCTTCACTGTGCAACTGTGCTGCTCACTGCTGCTTCTGATGCTTCGAT 120
Db      61  CTTGCTCAGAGCTTCACTGTGCAACTGTGCTGCTCACTGCTGCTTCTGATGCTTCGAT 120
QY      121  CCCAGAGGTTGCCCGGATCAGAGAGATTCCCTTGGAGAGGCTCTTCTGAGAA 180
Db      121  CCCAGAGGTTGCCCGGATCAGAGAGATTCCCTTGGAGAGGCTCTTCTGAGAA 180
QY      181  GATGACCACTGGGAGAGAGATCTGCCAGTAAAGATTCACCCAGAGAGAGAT 240
Db      181  GATGACCACTGGGAGAGAGATCTGCCAGTAAAGATTCACCCAGAGAGAGAT 240
QY      241  CCACCCGAGAGAGAGATCTACTGAGAGAGATCTACTGAGAGAGAGATCTACT 300
Db      241  CCACCCGAGAGAGAGATCTACTGAGAGAGATCTACTGAGAGAGAGATCTACT 300
QY      301  GAAATTAAAGCTTAATCAAGAAAGAGGCTCCTGAAATTAGAGATCTTACTACTGTT 360
Db      301  GAAATTAAAGCTTAATCAAGAAAGAGGCTCCTGAAATTAGAGATCTTACTACTGTT 360
QY      361  GAGGCTCTGAGAGATCTCAAGAAACCCAGAAATATCCACAGAGACAAAGAGGAT 420
Db      361  GAGGCTCTGAGAGATCTCAAGAAACCCAGAAATATCCACAGAGACAAAGAGGAT 420
QY      421  GACAGAGTCAATTGCGCTATGAGGCGACCCGCTGCGCCCGGCTGCCAGCCTGC 480
Db      421  GACAGAGTCAATTGCGCTATGAGGCGACCCGCTGCGCCCGGCTGCCAGCCTGC 480
QY      481  GCGGCGGCTTCCAGTCCCGGCTGATATCCGCGCCGAGCTCGCGGCTCTGCGCGGC 540
Db      481  GCGGCGGCTTCCAGTCCCGGCTGATATCCGCGCCGAGCTCGCGGCTCTGCGCGGC 540
QY      541  CTGCGCCCTGAACTCTCTGAGGCTTCAAGTCCCGGCTGCCAGAACTGCGCTGCGC 600
Db      541  CTGCGCCCTGAACTCTCTGAGGCTTCAAGTCCCGGCTGCCAGAACTGCGCTGCGC 600
QY      601  AACATGCGCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 660
Db      601  AACATGCGCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 660
QY      661  GGGCGGAGTACCGGCTCTGCACTGATCTGCACTGAGTCTGAGGAGGCTGAGGCT 720
Db      661  GGGCGGAGTACCGGCTCTGCACTGATCTGCACTGAGTCTGAGGAGGCTGAGGCT 720
QY      721  TCGAGACACTGTGGAAGGCAAGGCTTCCCTGCGAGATCCAGTGTTCACCTGAGC 780
Db      721  TCGAGACACTGTGGAAGGCAAGGCTTCCCTGCGAGATCCAGTGTTCACCTGAGC 780
QY      781  ACCGCTTTGCGAGAGTGAAGAGGCTTGGAGGCGCCGAGAGGCTGAGGCTGAGGCT 840
Db      781  ACCGCTTTGCGAGAGTGAAGAGGCTTGGAGGCGCCGAGAGGCTGAGGCTGAGGCT 840

```



```

Db      601 AACATGCGCCACATGTCGCACTGACCTTCCCTCCCTGAGCTAGAGATGCTCTGAGGTC 660
Qy      661 GGGGCGGAGTACCGGAGCTCTGCAAGCTGATCTGCACTGGGGGAGCTGCAAGTCTGCGGAC 720
Db      661 GGGGCGGAGTACCGGAGCTCTGCAAGCTGATCTGCACTGGGGGAGCTGCAAGTCTGCGGAC 720
Qy      721 TCGGAGGACACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
Db      721 TCGGAGGACACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
Qy      781 ACCGCTTTTGCAGAGTGAAGAGGCTTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
Db      781 ACCGCTTTTGCAGAGTGAAGAGGCTTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
Qy      841 GCCTTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
Db      841 GCCTTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
Qy      901 GAGGAAATCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
Db      901 GAGGAAATCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
Qy      961 CTGCGCTCTGACTGACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db      961 CTGCGCTCTGACTGACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Qy      1021 GCCGAGGAGTGTGATCTGAGCTGATGTTTAAACAGACAGTATGAGTGTGAGTGTGAGTGTGAG 1080
Db      1021 GCCGAGGAGTGTGATCTGAGCTGATGTTTAAACAGACAGTATGAGTGTGAGTGTGAGTGTGAG 1080
Qy      1081 CACACCTCTCTGACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
Db      1081 CACACCTCTCTGACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
Qy      1141 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
Db      1141 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
Qy      1201 AGTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
Db      1201 AGTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
Qy      1261 GCCCTGCTTTTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320
Db      1261 GCCCTGCTTTTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320
Qy      1321 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
Db      1321 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
Qy      1381 ACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
Db      1381 ACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
Qy      1441 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Db      1441 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Qy      1501 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1522
Db      1501 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1522

```

```

FILE REFERENCE: D-0021.5A
CURRENT APPLICATION NUMBER: US/09/177,776A
CURRENT FILING DATE: 1998-10-23
EARLIER APPLICATION NUMBER: 08/787,739
EARLIER FILING DATE: 1997-01-24
EARLIER APPLICATION NUMBER: 08/485,049
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 08/486,756
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 08/477,504
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 08/481,658
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 08/485,862
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 08/485,863
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 08/487,077
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 08/260,190
EARLIER FILING DATE: 1994-06-15
EARLIER APPLICATION NUMBER: 08/177,093
EARLIER FILING DATE: 1993-12-30
EARLIER APPLICATION NUMBER: 07/964,589
EARLIER FILING DATE: 1992-10-21
EARLIER APPLICATION NUMBER: PV-709-92
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 1
LENGTH: 1522
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: CDS
LOCATION: (13)..(1389)
FEATURE:
NAME/KEY: mat_peptide
LOCATION: (124)..(1389)
US-09-177-776-1
Query Match 100.0%; Score 1522; DB 3; Length 1522;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 ACAGTACGCGGATGAGCTCCCTGTGCCCCAGCCCTGGCTCCCTGTGATCCCGGCC 60
Db      1 ACAGTACGCGGATGAGCTCCCTGTGCCCCAGCCCTGGCTCCCTGTGATCCCGGCC 60
Qy      61 CCTGCTCCAGGAGCTCACTGTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Db      61 CCTGCTCCAGGAGCTCACTGTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Qy      121 CCCCAAGGTTGCCCGGATGCAAGAGATTTCCCTTGGAGAGAGGCTCTTCTGGGAA 180
Db      121 CCCCAAGGTTGCCCGGATGCAAGAGATTTCCCTTGGAGAGAGGCTCTTCTGGGAA 180
Qy      121 CCCCAAGGTTGCCCGGATGCAAGAGATTTCCCTTGGAGAGAGGCTCTTCTGGGAA 180
Db      121 CCCCAAGGTTGCCCGGATGCAAGAGATTTCCCTTGGAGAGAGGCTCTTCTGGGAA 180
Qy      181 GATGACCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Db      181 GATGACCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Qy      181 GATGACCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Db      181 GATGACCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Qy      241 CCAACCGGAGAGAGATCTACTGAGAGGAGATCTACTGAGAGGAGATCTACTGAGAGGAGATCTACT 300
Db      241 CCAACCGGAGAGAGATCTACTGAGAGGAGATCTACTGAGAGGAGATCTACTGAGAGGAGATCTACT 300
Qy      301 GAAGTTAAAGCTTAATTCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
Db      301 GAAGTTAAAGCTTAATTCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
Qy      361 GAGGCTCTGAGATCTCTCAAGAACCCCAAGATATGCCCACAGGAGCAAAAGAGGAGAT 420
Db      361 GAGGCTCTGAGATCTCTCAAGAACCCCAAGATATGCCCACAGGAGCAAAAGAGGAGAT 420

```

```

RESULT 10
US-09-177-776-1
; Sequence 1, Application US/09177776A
; Patent No. 6297051
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein

```



421 GACCAAGATATGCGCTATGAGGAGACCCGCTTGGCCCCGGGTGTCCCAAGCTTGC 480  
421 GACCAAGATATGCGCTATGAGGAGACCCGCTTGGCCCCGGGTGTCCCAAGCTTGC 480  
481 GCGGGGCGCTTCCAGTCCCGGATGATATCCGCCCCAGCTGGCGCTTGTGCGCGGCG 540  
481 GCGGGGCGCTTCCAGTCCCGGATGATATCCGCCCCAGCTGGCGCTTGTGCGCGGCG 540  
541 CTGCGCCCCCTGGAATCTCTGGGCTTCCAGCTTCCGCGCTTCCAGAACTGCGCTGCGC 600  
541 CTGCGCCCCCTGGAATCTCTGGGCTTCCAGCTTCCGCGCTTCCAGAACTGCGCTGCGC 600  
601 AACATGGCCCACTATGCAACTGACCTGCTCCTGGGCTAGATGCTTGGAGTCCC 660  
601 AACATGGCCCACTATGCAACTGACCTGCTCCTGGGCTAGATGCTTGGAGTCCC 660  
661 GGGCGGAGATACCGGGCTCTGACAGCTGATCTGACCTGAGGGGCTGACAGTGTCCGGCC 720  
661 GGGCGGAGATACCGGGCTCTGACAGCTGATCTGACCTGAGGGGCTGACAGTGTCCGGCC 720  
721 TCGAGAGACACTGTGGAAGGCAACCGCTTCCGCGAGATCCAGTGGTTCACCTCAGC 780  
721 TCGAGAGACACTGTGGAAGGCAACCGCTTCCGCGAGATCCAGTGGTTCACCTCAGC 780  
781 ACCGCTTTGGCAGAGTTGACGAGGCTTGGGGCGCCCGGAGGCTTGGCTGTGGGC 840  
781 ACCGCTTTGGCAGAGTTGACGAGGCTTGGGGCGCCCGGAGGCTTGGCTGTGGGC 840  
841 GCGTTTCTGAGAGAGGGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
841 GCGTTTCTGAGAGAGGGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
901 GAAGAAATCGCTGAG 960  
901 GAAGAAATCGCTGAG 960  
961 CTGCGCTTCTGACTTCAAGCGCTTCTTCAATATGAGGGGCTTCTGACTACCGGCTGT 1020  
961 CTGCGCTTCTGACTTCAAGCGCTTCTTCAATATGAGGGGCTTCTGACTACCGGCTGT 1020  
1021 GCCAGAGGTGCTCATCTGAGCTGTGATTTAACAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
1021 GCCAGAGGTGCTCATCTGAGCTGTGATTTAACAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
1081 CACACCTCTCTGACACCTGTGTGGGAGCTGTGAGCTCTGCTGAGCTGAACTTCCGA 1140  
1081 CACACCTCTCTGACACCTGTGTGGGAGCTGTGAGCTCTGCTGAGCTGAACTTCCGA 1140  
1141 GCGAGGAGAGCTTTGAAATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
1141 GCGAGGAGAGCTTTGAAATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
1201 AGTCTCGGGCTGTGAG 1260  
1201 AGTCTCGGGCTGTGAG 1260  
1261 GCGCTGCTTTGGGCTGCTTTTGGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
1261 GCGCTGCTTTGGGCTGCTTTTGGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
1321 AGGAG 1380  
1321 AGGAG 1380  
1381 ACTGAGAGCTGAG 1440  
1381 ACTGAGAGCTTGAAG 1440  
1441 GAG 1500  
1441 GAG 1500  
1501 TTTTAAATTAATTTTAAAT 1522

Db 1501 TTTTAAATTAATTTTAAAT 1522

RESULT 11  
US-08-335-469-1  
Sequence 1, Application US/08335469A  
Patent No. 604535  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: Methods to Detect and Quantify MN Protein/Polypeptide  
TITLE OF INVENTION: Using MN-Specific Antibodies  
FILE REFERENCE: D-0021A  
CURRENT APPLICATION NUMBER: US/08/335,469A  
CURRENT FILING DATE: 1994-11-07  
EARLIER APPLICATION NUMBER: 07/964,589  
EARLIER FILING DATE: 1992-10-21  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 1399  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1266)  
US-08-335-469-1

Query Match 91.9%; Score 1399; DB 3; Length 1399;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

124 CAGAGGTGCCCCGAGATGAGAGAGATTCCTTGGAGAGAGAGCTTCTTGGGAGAGAT 183  
1 CAGAGGTGCCCCGAGATGAGAGAGATTCCTTGGAGAGAGAGCTTCTTGGGAGAGAT 60  
184 GACCACTGGGCGAGAGAGATTCGCCAGTGAAGAGATTCACCCAGAGAGAGAGATCA 243  
61 GACCACTGGGCGAGAGAGATTCGCCAGTGAAGAGATTCACCCAGAGAGAGAGATCA 120  
244 CCGGAGAGAGAGATTCCTGAGAGAGAGATTCACCTGAGAGAGAGAGATTCACCTGA 303  
121 CCGGAGAGAGAGATTCCTGAGAGAGAGATTCACCTGAGAGAGAGAGATTCACCTGA 180  
304 GTTAAGCTTAATCAG 363  
181 GTTAAGCTTAATCAG 240  
364 GTCCTGAGAGATCTCAAGACCCAGAGATTAAGCCAGAGAGAGAGAGAGAGAGAGAG 423  
241 GTCCTGAGAGATCTCAAGACCCAGAGATTAAGCCAGAGAGAGAGAGAGAGAGAGAG 300  
424 CAGAGTATTGGGCTATGAG 483  
301 CAGAGTATTGGGCTATGAG 360  
484 GGGCGCTTCAAGTCCCGGAGATATCCGCCCCAGAGCTCGCGCTTGTGCGCGGCTG 543  
361 GGGCGCTTCAAGTCCCGGAGATATCCGCCCCAGAGCTCGCGCTTGTGCGCGGCTG 420  
544 CGCCCCCTGGAATCTCTGGGCTTCCAGCTCCGCGCTTCCAGAACTGCGCTGCGCAAC 603  
421 CGCCCCCTGGAATCTCTGGGCTTCCAGCTCCGCGCTTCCAGAACTGCGCTGCGCAAC 480  
604 AATGGCCAGAGTGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663  
481 AATGGCCAGAGTGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
664 CCGAGATACCGGAGCTTGGAGCTGAGCTGCACTGAGAGAGAGAGAGAGAGAGAGAGAG 723  
541 CCGAGATACCGGAGCTTGGAGCTGAGCTGCACTGAGAGAGAGAGAGAGAGAGAGAGAG 600



QY 663 GCGGAGTACCGGCTCTGACGTGCTGCTGAGGAGGCTGCTGCGGCTC 722  
DB 538 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 596  
QY 723 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 782  
DB 597 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 656  
QY 783 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 842  
DB 657 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 715  
QY 843 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 901  
DB 716 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 775  
QY 902 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 961  
DB 776 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 835  
QY 962 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1021  
DB 836 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 895  
QY 1022 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1081  
DB 896 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 955  
QY 1082 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1141  
DB 956 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1015  
QY 1142 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1201  
DB 1016 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1075  
QY 1202 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1261  
DB 1076 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1135  
QY 1262 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1321  
DB 1136 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1195  
QY 1322 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1380  
DB 1196 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1255  
QY 1381 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1440  
DB 1256 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1315  
QY 1441 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1500  
DB 1316 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1375  
QY 1501 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1562  
DB 1376 GCGGAGTACCGGCTCTGACGTGCTGAGGAGGCTGCTGCGGCTC 1397

## RESULT 13

PCT-US93-02024-1

Sequence 1, Application PC/TUS9302024  
GENERAL INFORMATION:  
APPLICANT: Ciba Corning Diagnostics Corp.  
TITLE OF INVENTION: Institute of Virology  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: Stewart Street Tower, 18th Fl., One Market  
STREET: Plaza

CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/02024  
FILING DATE: 19930308  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: CS PV-709-92  
FILING DATE: 10-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/964,589  
FILING DATE: 21-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-777-9257  
TELEFAX: 415-543-4219  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1397 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1287  
PCT-US93-02024-1  
Query Match 85.6%; Score 1302.8; DB 5; Length 1397;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1992; Conservative 0; Mismatches 2; Indels 8; Gaps 8;  
QY 124 CAGAGTGTGCCCCGAGTGCAGAGGATTCCTGAGAGGAGGCTCTTGGGAGAT 183  
DB 1 CAGAGTGTGCCCCGAGTGCAGAGGATTCCTGAGAGGAGGATTCCTGAGAGAT 59  
QY 184 GACCCACTGGGCGAGAGGATTCCTGAGAGGAGGATTCCTGAGAGGAGATTCGA 243  
DB 60 GACCCACTGGGCGAGAGGATTCCTGAGAGGAGGATTCCTGAGAGGAGATTCGA 119  
QY 244 CCGGAGAGGAGATTCCTGAGAGGAGGATTCCTGAGAGGAGATTCCTGAGAA 303  
DB 120 CCGGAGAGGAGATTCCTGAGAGGAGGATTCCTGAGAGGAGATTCCTGAGAA 179  
QY 304 GTTA-GGCTAAATCAAGAGAGGAGGCTCCGAGGATTCCTGAGAGGATTCCTGAG 362  
DB 180 GTTAATGCTTAATCAAGAGAGGAGGCTCCGAGGATTCCTGAGAGGATTCCTGAG 239  
QY 363 GAGTCTGAGATTCCTGAGAGGAGGATTCCTGAGAGGAGGATTCCTGAGAGGATTCGA 422  
DB 240 GAGTCTGAGATTCCTGAGAGGAGGATTCCTGAGAGGAGGATTCCTGAGAGGATTCGA 299  
QY 423 CAGAGTATTCCTGAGAGGAGGATTCCTGAGAGGAGGATTCCTGAGAGGATTCCTGAG 482  
DB 300 CAGAGTATTCCTGAGAGGAGGATTCCTGAGAGGAGGATTCCTGAGAGGATTCCTGAG 358  
QY 483 GAGGCGCTTCAGTCCCGGAGGATTCCTGAGAGGAGGATTCCTGAGAGGATTCCTGAG 542  
DB 359 GAGGCGCTTCAGTCCCGGAGGATTCCTGAGAGGAGGATTCCTGAGAGGATTCCTGAG 418  
QY 543 GAGGCGCTTCAGTCCCGGAGGATTCCTGAGAGGAGGATTCCTGAGAGGATTCCTGAG 602

419 GCGCCCTGGAACTCTTGGGCTTCCAGCTCCCGCGCTCCCGAACTGGCGCTGC-AGA 477  
QY 603 CAATGGCCACAGTGTGCACTGACCTGCTCTTGGGCTTGAAGATGGCTCTGGGTCGGG 662  
Db 478 CAATGGCCACAGTGTGCACTGACCTGCTCTTGGGCTTGAAGATGGCTCTGGGTCGGG 537  
QY 663 GCGGAGTACCGGGCTCTGACAGTGTGCACTGGGCGGCTGACAGTGTCTCGGGGCTC 722  
Db 538 GCGGAGTACCGGGCTCTGACAGTGTGCACTGGGCGGCTGACAGTGTCTCGGGGCTC 596  
QY 723 GAGACACACTGTGGAAGGCCACCGCTTCCCTGCGAGATCAGCTGCTTACCTGACAC 782  
Db 597 GAGACACACTGTGGAAGGCCACCGCTTCCCTGCGAGATCAGCTGCTTACCTGACAC 656  
QY 783 GCGCTTGGCCAGATTTGACAGAGGCTTGGGCGGCTGAGGCGCTGGCGGCTTGGCGG 842  
Db 657 GCGCTTGGCCAGATTTGACAGAGGCTTGGGCGGCTGAGGCGCTGGCGGCTTGGCGG 715  
QY 843 CTTTCTGAGAGAGGCGCGGAGAGAAACAATG-CCATGAGCAGTGTCTGCTGCTTGG 901  
Db 716 CTTTCTGAGAGAGGCGCGGAGAGAAACAATG-CCATGAGCAGTGTCTGCTGCTTGG 775  
QY 902 AAGAATGCTGAGAGAGGCTGACAGACTAGGCTCCAGAGCTGACATATCTGACATTC 961  
Db 776 AAGAATGCTGAGAGAGGCTGACAGACTAGGCTCCAGAGCTGACATATCTGACATTC 835  
QY 962 TGCCCTCTGACTGAGCGGCTACTTCCATATAGAGGAGTCTCTGACTACAGCGGCTGTG 1021  
Db 836 TGCCCTCTGACTGAGCGGCTACTTCCATATAGAGGAGTCTCTGACTACAGCGGCTGTG 895  
QY 1022 CCCAGGCTGCTGAGCTGTGTTTAAACAAGAGTGTGCTGAGTGTCTTAAAGCTTC 1081  
Db 896 CCCAGGCTGCTGAGCTGTGTTTAAACAAGAGTGTGCTGAGTGTCTTAAAGCTTC 955  
QY 1082 ACACCTCTCTGACACCTGTGGGAGACTGTGACTCTGCGGCTACAGCTGAACTTCCAG 1141  
Db 956 ACACCTCTCTGACACCTGTGGGAGACTGTGACTCTGCGGCTACAGCTGAACTTCCAG 1015  
QY 1142 CGACGACAGCTTGAATGAGCGAGATGTTAGGCTCTTCCCTGCTGAGTGAAGCA 1201  
Db 1016 CGACGACAGCTTGAATGAGCGAGATGTTAGGCTCTTCCCTGCTGAGTGAAGCA 1075  
QY 1202 GTCCTGGGCGTCTGAGCCAGTCCAGTGAATTCGCGCGGCTGTGCTGATCCCTAG 1261  
Db 1076 GTCCTGGGCGTCTGAGCCAGTCCAGTGAATTCGCGCGGCTGTGCTGATCCCTAG 1135  
QY 1262 CCGTGGTCTTGGGCTCTTCTTGTGCTGACACAGCGTGGGCTCTTGTGAGATGAGA 1321  
Db 1136 CCGTGGTCTTGGGCTCTTCTTGTGCTGACACAGCGTGGGCTCTTGTGAGATGAGA 1195  
QY 1322 GCGAGCACAAGAGGGAACCAAGAGGAGTGTAGC-TACCGCCACAGAGATTAACGAG 1380  
Db 1196 GCGAGCACAAGAGGGAACCAAGAGGAGTGTAGC-TACCGCCACAGAGATTAACGAG 1255  
QY 1381 ACTGAGACCTAGAGGCTGATCTTGTGGAATGTGAGAAGCCAGCCAGAGCATCTGAGG 1440  
Db 1256 ACTGAGACCTAGAGGCTGATCTTGTGGAATGTGAGAAGCCAGCCAGAGCATCTGAGG 1315  
QY 1441 GGAAGCCGCTACTCTCTGCTGCTGCTCATTTAGCCACTTCTTAACTGCAAGAAAT 1500  
Db 1316 GGAAGCCGCTACTCTCTGCTGCTGCTCATTTAGCCACTTCTTAACTGCAAGAAAT 1375  
QY 1501 TTTTAAATTAATTTTATAT 1522  
Db 1376 TTTTAAATTAATTTTATAT 1397

RESULT 14  
US-08-481-658B-5  
; Sequence 5, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:

APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
City: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30 (EPC)  
CURRENT APPLICATION DATA:  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021,3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10898 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-481-658B-5  
Query Match 27.2%; Score 414.4; DB 2; Length 10898;  
Best Local Similarity 99.8%; Pred. No. 7.4e-94;  
Matches 415; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ACAGTACCGCGATGAGCTCCCTGCTGCGCCAGCCCTTGGCTCTGTTATCCGAGC 60  
Db 3537 ACAGTACCGCGATGAGCTCCCTGCTGCGCCAGCCCTTGGCTCTGTTATCCGAGC 3596  
QY 61 CCGTCTCCAGGCGCTCACTGTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
Db 3597 CCGTCTCCAGGCGCTCACTGTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3656  
QY 121 CCCAGAGGTTGCCCGGATGCAAGAGATTCGCCCTTGGAGAGAGCTCTTGGGAA 180  
Db 3657 CCCAGAGGTTGCCCGGATGCAAGAGATTCGCCCTTGGAGAGAGCTCTTGGGAA 3716  
QY 181 GATGACCCACTGGGAGAGAGATTCGCCAGTGAAGAGATTACCCAGAGAGAGAT 240  
Db 3717 GATGACCCACTGGGAGAGAGATTCGCCAGTGAAGAGATTACCCAGAGAGAGAT 3776  
QY 241 CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTTACCTGAGAGAGATCTACCT 300  
Db 3777 CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTTACCTGAGAGAGATCTACCT 3836  
QY 301 GAAGTTAAGCTTAATTCAGAGAGAGGCTCCCTGAGAGATTAAGAGATCTACCT 360  
Db 3837 GAAGTTAAGCTTAATTCAGAGAGAGGCTCCCTGAGAGATTAAGAGATCTACCT 3896  
QY 361 GAGGCTCTGAGAGATCTCAAGAACCCAGATATATGCCCACAGAGAGAGAGAG 416  
Db 3897 GAGGCTCTGAGAGATCTCAAGAACCCAGATATATGCCCACAGAGAGAGAGAG 3952





INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1522 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-772-713-1

Query Match 100.0%; Score 1522; DB 9; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTACGCGGATGGCTCCCTGTCGCCAGCCGCTGGCTCCCTGTTGATCCGCGCC 60  
DB 1 AAGTACGCGGATGGCTCCCTGTCGCCAGCCGCTGGCTCCCTGTTGATCCGCGCC 60  
QY 61 CCTGCTCCAGGCTCTCACTGTGCACTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
DB 61 CCTGCTCCAGGCTCTCACTGTGCACTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
QY 121 CCCGAGAGGTTTCCCCGATGAGAGAGATTCGCCCTTGGAGAGAGGCTTTCTGGGGA 180  
DB 121 CCCGAGAGGTTTCCCCGATGAGAGAGATTCGCCCTTGGAGAGAGGCTTTCTGGGGA 180  
QY 181 GATGACCCACTGGGAGAGAGATTCGCCCTGAGAGAGATTCACCACAGAGAGAGAT 240  
DB 181 GATGACCCACTGGGAGAGAGATTCGCCCTGAGAGAGATTCACCACAGAGAGAGAT 240  
QY 241 CCACCCGAGAGAGAGATTCCTACTGAGAGAGATTCCTGAGAGAGAGATTCCTACT 300  
DB 241 CCACCCGAGAGAGAGATTCCTACTGAGAGAGATTCCTGAGAGAGAGATTCCTACT 300  
QY 301 GAAGTTAAGCTTAATCAGAAAGAGAGGCTCCCTGAAGTTAGAGATTCCTACTGTT 360  
DB 301 GAAGTTAAGCTTAATCAGAAAGAGAGGCTCCCTGAAGTTAGAGATTCCTACTGTT 360  
QY 361 GAGGCTCTGAGAGATTCCTCAAGAACCCCAAGATTAATGCCCAAGAGAGAGAGAT 420  
DB 361 GAGGCTCTGAGAGATTCCTCAAGAACCCCAAGATTAATGCCCAAGAGAGAGAGAT 420  
QY 421 GACCAAGATTCATGGCGCTATGAGAGAGAGCCGCGCTGGCCGGGCTGTCCAGCTGC 480  
DB 421 GACCAAGATTCATGGCGCTATGAGAGAGAGCCGCGCTGGCCGGGCTGTCCAGCTGC 480  
QY 481 GCGGCGCGCTTCAGTCCCGGTGAGATTCGCGCCCAAGCTGCGCGCTTCTGCGCGCC 540  
DB 481 GCGGCGCGCTTCAGTCCCGGTGAGATTCGCGCCCAAGCTGCGCGCTTCTGCGCGCC 540  
QY 541 CTGGCGCCCTGGAATCCTGAGGCTTCGAGCTCCCGCGCTCCCAAGATTCGCGCGCC 600  
DB 541 CTGGCGCCCTGGAATCCTGAGGCTTCGAGCTCCCGCGCTCCCAAGATTCGCGCGCC 600  
QY 601 AACCAATGGCCACAGTGTGCACTGACCTGCTCTGAGGCTAGAGATGAGTGTGCTGCC 660  
DB 601 AACCAATGGCCACAGTGTGCACTGACCTGCTCTGAGGCTAGAGATGAGTGTGCTGCC 660  
QY 661 GGGCGGAGATACCGGCTCTGAGCTGATCTGACATGAGGAGGCTGAGAGTCCGCGCG 720  
DB 661 GGGCGGAGATACCGGCTCTGAGCTGATCTGACATGAGGAGGCTGAGAGTCCGCGCG 720  
QY 721 TCGAGACACACTGTGGAAGGCCACCGTTCCCTGCGAGATTCACGTTGATCACTGAGC 780  
DB 721 TCGAGACACACTGTGGAAGGCCACCGTTCCCTGCGAGATTCACGTTGATCACTGAGC 780  
QY 781 ACCGCTTTGCAAGATTGACAGAGCTTTGGGAGGCGCCGGAGAGCTGGCGCTGTGGCC 840  
DB 781 ACCGCTTTGCAAGATTGACAGAGCTTTGGGAGGCGCCGGAGAGCTGGCGCTGTGGCC 840  
QY 841 GCCTTTCTGAGAGAGGCGCCGGAAGAAAACAGTGGCTTATGACAGTTGCTGTCTGCTTG 900

DB 841 GCCTTTCTGAGAGAGGCGCCGGAAGAAAACAGTGGCTTATGACAGTTGCTGTCTGCTTG 900  
QY 901 GAAGAAATCGGTGAGAGAGGCTCAGAGACTCAGTCTCCCGAGATGAGATATCTGACTC 960  
DB 901 GAAGAAATCGGTGAGAGAGGCTCAGAGACTCAGTCTCCCGAGATGAGATATCTGACTC 960  
QY 961 CTGCGCTCTGACTCAGAGGCTCTCTCAATATGAGAGGCTCTCTGACTACACCGGCTGT 1020  
DB 961 CTGCGCTCTGACTCAGAGGCTCTCTCAATATGAGAGGCTCTCTGACTACACCGGCTGT 1020  
QY 1021 GCCCAGGCTGATCTGAGATCTGTGTTTAAACAGACAGTGAATGCTGAGTCTTAAGACTTC 1080  
DB 1021 GCCCAGGCTGATCTGAGATCTGTGTTTAAACAGACAGTGAATGCTGAGTCTTAAGACTTC 1080  
QY 1081 CACACCTCTCTGACACCTCTGTGGGAGCTGTGAGATCTCTGAGTACAGTGAATCTTCCGA 1140  
DB 1081 CACACCTCTCTGACACCTCTGTGGGAGCTGTGAGATCTCTGAGTACAGTGAATCTTCCGA 1140  
QY 1141 GCGACGACGCTTTGAATGAGGCGAGATGATGAGGCTCTCTCTGCTGAGTGAACAGC 1200  
DB 1141 GCGACGACGCTTTGAATGAGGCGAGATGATGAGGCTCTCTCTGCTGAGTGAACAGC 1200  
QY 1201 AGTCTGAGGCTGTGAGAGCCAGTCCAGTGAATTCCTGCTGAGTGTGAGATCTTA 1260  
DB 1201 AGTCTGAGGCTGTGAGAGCCAGTCCAGTGAATTCCTGCTGAGTGTGAGATCTTA 1260  
QY 1261 GCGCTGATTTTGGCTCTCTTTTGTGTCACAGAGGCTGCGCTCTTGTGAGATGAGA 1320  
DB 1261 GCGCTGATTTTGGCTCTCTTTTGTGTCACAGAGGCTGCGCTCTTGTGAGATGAGA 1320  
QY 1321 AGGACGACACGAGAGGAGACCAAGAGGAGTGTAGTACCGCCACAGACAGGATGCGAG 1380  
DB 1321 AGGACGACACGAGAGGAGACCAAGAGGAGTGTAGTACCGCCACAGACAGGATGCGAG 1380  
QY 1381 ACTGAGCTTGAAGGCTGTATCTTGGAGATGTGAGAACCAAGCAGAGACATCTAGGG 1440  
DB 1381 ACTGAGCTTGAAGGCTGTATCTTGGAGATGTGAGAACCAAGCAGAGACATCTAGGG 1440  
QY 1441 GAGCGGCTAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
DB 1441 GAGCGGCTAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
QY 1501 TTTTAAATTAATTTTAAAT 1522  
DB 1501 TTTTAAATTAATTTTAAAT 1522

RESULT 2  
US-09-967-237-1  
; Sequence 1, Application US/09967237  
; Publication No. US20030049828A1  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; TITLE OF INVENTION: MN Gene and Protein  
; FILE REFERENCE: D-0021.SB-2  
; CURRENT APPLICATION NUMBER: US/09/967,237  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR FILING DATE: 1998-10-23  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 1  
; LENGTH: 1522  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (13)..(1389)  
; NAME/KEY: mat\_peptide  
; LOCATION: (124)..(1389)  
US-09-967-237-1



Query Match 100.0%; Score 1522; DB 10; Length 1522;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AAGAGCAGCCGAGTGGCTCCCTGTCGCCCCAGCCCTGGCTCCCTGCTGTTGATCCCGGCC 60
DB 1 AAGAGCAGCCGAGTGGCTCCCTGTCGCCCCAGCCCTGGCTCCCTGCTGTTGATCCCGGCC 60
QY CCTGCTCAAGGCTCACTGTGCACTGCTGCTCACTGCTGCTTCTGATGCTGTCAT 120
DB CCTGCTCAAGGCTCACTGTGCACTGCTGCTCACTGCTGCTTCTGATGCTGTCAT 120
QY 121 CCCCAGAGTTGCCCCGATGTCAGAGATTTCCCTTTGGAGAGAGCTCTTTGGGGAA 180
DB 121 CCCCAGAGTTGCCCCGATGTCAGAGATTTCCCTTTGGAGAGAGCTCTTTGGGGAA 180
QY 181 GATGACCCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT 240
DB 181 GATGACCCACTGGGCGAGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT 240
QY 241 CCACCCGAGAGAGAGATCTTACCTGGAAGAGATCTTACCTGGAAGAGAGATCTTACT 300
DB 241 CCACCCGAGAGAGAGATCTTACCTGGAAGAGATCTTACCTGGAAGAGAGATCTTACT 300
QY 301 GAAGTTAAGCCTAAATCAGAGAGAGAGGCTCCCTGAAGTTAGAGATCTTACTACTGTT 360
DB 301 GAAGTTAAGCCTAAATCAGAGAGAGAGGCTCCCTGAAGTTAGAGATCTTACTACTGTT 360
QY 361 GAGGCTCTGAGATCTTCAAGAACCCAGAAATATGCCCAAGAGGAAAGAGAGAT 420
DB 361 GAGGCTCTGAGATCTTCAAGAACCCAGAAATATGCCCAAGAGGAAAGAGAGAT 420
QY 421 GACCCAGAGATCTGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 GACCCAGAGATCTGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 GGGGCGCGCTTCAGTCCCGGGTGAATATCCGCCCCAGACTGCGCGCTTTGCGCGGCC 540
DB 481 GGGGCGCGCTTCAGTCCCGGGTGAATATCCGCCCCAGACTGCGCGCTTTGCGCGGCC 540
QY 541 CTGCGCCCCCTGGAATCTTGCGCTTCCAGCTCCCGCGCTCCCAAGAACTGCGCTGCG 600
DB 541 CTGCGCCCCCTGGAATCTTGCGCTTCCAGCTCCCGCGCTCCCAAGAACTGCGCTGCG 600
QY 601 AACATGCGCACAGTGTGCACTGACCTGCTCTGCGGCTTGAAGATGCTCTGCGTCCC 660
DB 601 AACATGCGCACAGTGTGCACTGACCTGCTCTGCGGCTTGAAGATGCTCTGCGTCCC 660
QY 661 GGGCGGAGTACCGGGCTCTGCACTGCACTGCGGGGCTGCAAGTGTGTCGGGCC 720
DB 661 GGGCGGAGTACCGGGCTCTGCACTGCACTGCGGGGCTGCAAGTGTGTCGGGCC 720
QY 721 TCGGAGCACACTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 TCGGAGCACACTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 ACCGCTTTGCCAGAGATTGACAGAGGCTTGGGGGCGCCGGAGAGGCTGTGGCTGTGGCC 840
DB 781 ACCGCTTTGCCAGAGATTGACAGAGGCTTGGGGGCGCCGGAGAGGCTGTGGCTGTGGCC 840
QY 841 GCGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 GCGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 GAAGAATCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 GAAGAATCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 961 CTGCGCTCTGACTTACCGCGCTTCAATATGAGAGAGGCTCTGACTACACCGCCCTGT 1020
DB 961 CTGCGCTCTGACTTACCGCGCTTCAATATGAGAGAGGCTCTGACTACACCGCCCTGT 1020

```

```

QY 1021 GCCAGAGGTGTCATCTGACTGTGTGTTTAAACCAAGACAGTGTGCTGAGTGAAGAGCTC 1080
DB 1021 GCCAGAGGTGTCATCTGACTGTGTGTTTAAACCAAGACAGTGTGCTGAGTGAAGAGCTC 1080
QY 1081 CACACCCCTCTGACACCCCTGTGGAGAGCTGTGACTCTGAGCTGAGCTGAAGCTTCCGA 1140
DB 1081 CACACCCCTCTGACACCCCTGTGGAGAGCTGTGACTCTGAGCTGAGCTGAAGCTTCCGA 1140
QY 1141 GCGAGCAGGCTTTGAAATGGGCGAGTATGAGGCTCTTCCCTGCTGAGTGAAGAGC 1200
DB 1141 GCGAGCAGGCTTTGAAATGGGCGAGTATGAGGCTCTTCCCTGCTGAGTGAAGAGC 1200
QY 1201 AGTCTGGGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 AGTCTGGGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 GCGCTGCTTTTGGCTCTTCTTCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 GCGCTGCTTTTGGCTCTTCTTCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 AGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 AGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 ACTGAGGCTGAGAGCTGATCTTGAAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 ACTGAGGCTGAGAGCTGATCTTGAAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 GAGGCGGTATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 GAGGCGGTATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 TTTTAAATTAATTTTAAAT 1522
DB 1501 TTTTAAATTAATTTTAAAT 1522

RESULT 3
US-09-954-456-89
; Sequence 89, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954, 456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233, 617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234, 052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234, 923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235, 134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235, 637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235, 638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235, 711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 89
; LENGTH: 1552
; TYPE: DNA

```

ORGANISM: Homo sapiens  
US-09-954-456-89

Query Match 100.0%; Score 1522; DB 9; Length 1552;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ACAGTCAGCCGAGTGGCTCCCTGCTGCTCCGAGCCCTGCTCCCTGCTGTTATCCCGGAC 60
DB 31 ACAGTCAGCCGAGTGGCTCCCTGCTGCTCCGAGCCCTGCTCCCTGCTGTTATCCCGGAC 90
QY 61 CCTGCTCCAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
DB 91 CCTGCTCCAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 150
QY 121 CCCGAGAGTGGCCCGGATGCAAGAGATTCCTCCCTTGGAGAGAGGCTCTTCTGGGAA 180
DB 151 CCCGAGAGTGGCCCGGATGCAAGAGATTCCTCCCTTGGAGAGAGGCTCTTCTGGGAA 210
QY 181 GATGACCCACTGGGCGAGAGAGATTCCTCCAGTGAAGAGATTCACCCAGAGAGAGAT 240
DB 211 GATGACCCACTGGGCGAGAGAGATTCCTCCAGTGAAGAGATTCACCCAGAGAGAGAT 270
QY 241 CCACCCGAGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 300
DB 271 CCACCCGAGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 330
QY 301 GAAGTTAAGCTTAATCAGAAAGAGAGGCTCCCTTAAGTTAAGAGATCTACTACT 360
DB 331 GAAGTTAAGCTTAATCAGAAAGAGAGGCTCCCTTAAGTTAAGAGATCTACTACT 390
QY 361 GAGGCTCTGAGAGATCTCAAGAAACCCAGAAATATCCCAAGAGAGAGAGAGAGAT 420
DB 391 GAGGCTCTGAGAGATCTCAAGAAACCCAGAAATATCCCAAGAGAGAGAGAGAT 450
QY 421 GACGAGATCAATGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480
DB 451 GACGAGATCAATGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 510
QY 481 GGGGGCGGCTTCAGATCCCGGAGATATCCGAGCCGAGCTGCGCGCTCTGCTCCGAGC 540
DB 511 GGGGGCGGCTTCAGATCCCGGAGATATCCGAGCCGAGCTGCGCGCTCTGCTCCGAGC 570
QY 541 CTGCGCCCTTGGAACTCTGAGGCTTCCAGCTCCGCGCTCCCAAGATCGGCTGCGC 600
DB 571 CTGCGCCCTTGGAACTCTGAGGCTTCCAGCTCCGCGCTCCCAAGATCGGCTGCGC 630
QY 601 AACATGGCCACAGTGCACATCTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 631 AACATGGCCACAGTGCACATCTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 690
QY 661 GGGCGGAGTACCGGCTCTGAGTGCATCTGCACTGGGAGGCTGCAAGTCTGCGGAGC 720
DB 691 GGGCGGAGTACCGGCTCTGAGTGCATCTGCACTGGGAGGCTGCAAGTCTGCGGAGC 750
QY 721 TGGAGACACTGTGGAAGGCAACGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 751 TGGAGACACTGTGGAAGGCAACGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 810
QY 781 ACCGCTTGGCAGAGTTCAGAGGCTTGGGAGGCTGGGAGGCTGCGCTGCTGCTGCTGCT 840
DB 811 ACCGCTTGGCAGAGTTCAGAGGCTTGGGAGGCTGGGAGGCTGCGCTGCTGCTGCTGCT 870
QY 841 GCTTTTGGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 871 GCTTTTGGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 930
QY 901 GAGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 931 GAGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 990
QY 961 CTGCTCTGACTTCAGCGGCTACTTCCAAATATGAGGAGTCTGATCTAACCGGCTCTG 1020
```

```

DB 991 CTGCTCTGACTTCAGCGGCTACTTCCAAATATGAGGAGTCTGATCTAACCGGCTCTG 1050
QY 1021 GCCCAGAGGTGATCTGAGCTGTGTTTAAACGACAGTATGATGATGCTTAAGCAGCTC 1080
DB 1051 GCCCAGAGGTGATCTGAGCTGTGTTTAAACGACAGTATGATGATGCTTAAGCAGCTC 1110
QY 1081 CACACCTCTCTGACACCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1111 CACACCTCTCTGACACCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1170
QY 1141 GCGACGAGCTTTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1171 GCGACGAGCTTTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1230
QY 1201 AGTCTCTGAGGCTGCTGAGCAGTCCAGCTGAATTCCTGCTGAGCTGCTGAGATCTCTA 1260
DB 1231 AGTCTCTGAGGCTGCTGAGCAGTCCAGCTGAATTCCTGCTGAGCTGCTGAGATCTCTA 1290
QY 1261 GCCCTGATTTTGGCTCTCTTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
DB 1291 GCCCTGATTTTGGCTCTCTTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1350
QY 1321 AGGACGACAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1351 AGGACGACAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1410
QY 1381 ACTGAGGCTTGAAGAGTCTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1411 ACTGAGGCTTGAAGAGTCTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1470
QY 1441 GAGCGGCTTGAAGAGTCTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1471 GAGCGGCTTGAAGAGTCTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1530
QY 1501 TTTTAAATTAATTAATTAAT 1522
DB 1531 TTTTAAATTAATTAATTAAT 1552
```

```

RESULT 4
US-09-954-456-726
; Sequence 726, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 726
```



QY 61 CTTGCTCCAGGCTCACTGTGCACTGTGCTGTCACTGTGCTTCTGATGCTGTCCAT 120  
 Db 91 CTTGCTCCAGGCTCACTGTGCACTGTGCTGTCACTGTGCTTCTGATGCTGTCCAT 150  
 QY 121 CCCAGAGGTTGCCCCGAGATGCAAGAGATTTCCCCCTTGGAGAGAGCTCTTCTGGGAA 180  
 Db 151 CCCAGAGGTTGCCCCGAGATGCAAGAGATTTCCCCCTTGGAGAGAGCTCTTCTGGGAA 210  
 QY 181 GATGATCCACTGGGCGAGAGATCTGCCAAGTGAAGAGATTTCAACCAAGAGAGAGAT 240  
 Db 211 GATGATCCACTGGGCGAGAGATCTGCCAAGTGAAGAGATTTCAACCAAGAGAGAGAT 270  
 QY 241 CCAACCGGAGAGAGATCTAAGAGAGAGATCTAAGAGAGAGATCTAAGAGAGAGATCTAAG 300  
 Db 271 CCAACCGGAGAGAGATCTAAGAGAGAGATCTAAGAGAGAGATCTAAGAGAGAGATCTAAG 330  
 QY 301 GAAATTAGCTTAATCAGAT 360  
 Db 331 GAAATTAGCTTAATCAGAT 390  
 QY 361 GAGGCTCTGAGAGATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 420  
 Db 391 GAGGCTCTGAGAGATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 450  
 QY 421 GAGGCTCTGAGAGATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480  
 Db 451 GAGGCTCTGAGAGATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 510  
 QY 481 GGGGGGGGCTTCCAGTCCCCGGGTGATATCCGGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
 Db 511 GGGGGGGGCTTCCAGTCCCCGGGTGATATCCGGGCGGCGGCGGCGGCGGCGGCGGCGG 570  
 QY 541 CTGCGCCCTTGGAGTCTTGGAGTCTTGGAGTCTTGGAGTCTTGGAGTCTTGGAGTCTTGG 600  
 Db 571 CTGCGCCCTTGGAGTCTTGGAGTCTTGGAGTCTTGGAGTCTTGGAGTCTTGGAGTCTTGG 630  
 QY 601 AACAAATGGCCACAGTGTGCACTGACCTTGTCTTGGAGTCTTGGAGTCTTGGAGTCTTGG 660  
 Db 631 AACAAATGGCCACAGTGTGCACTGACCTTGTCTTGGAGTCTTGGAGTCTTGGAGTCTTGG 690  
 QY 661 GGGGGGAGTACCGGGCTGTGAGTGTGACCTGTGACCTGTGAGTGTGAGTGTGAGTGTGAG 720  
 Db 691 GGGGGGAGTACCGGGCTGTGAGTGTGACCTGTGACCTGTGAGTGTGAGTGTGAGTGTGAG 750  
 QY 721 TCGAGACACACTGTGAT 780  
 Db 751 TCGAGACACACTGTGAT 810  
 QY 781 ACCGCTTTTGCAGAGTGTGAT 840  
 Db 811 ACCGCTTTTGCAGAGTGTGAT 870  
 QY 841 GCTTTTCTGAT 900  
 Db 871 GCTTTTCTGAT 930  
 QY 901 GAAGAATCGCTAGAT 960  
 Db 931 GAAGAATCGCTAGAT 990  
 QY 961 CTGCTCTGACTTCAAGCGGCTAATTCACAAATAGAGGAGTCTGACTTCAAGCGGCTGT 1020  
 Db 991 CTGCTCTGACTTCAAGCGGCTAATTCACAAATAGAGGAGTCTGACTTCAAGCGGCTGT 1050  
 QY 1021 GCCCAGAGGAGTCACTGAT 1080  
 Db 1051 GCCCAGAGGAGTCACTGAT 1110  
 QY 1081 CACACCTCTCTGACACCTCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140  
 Db 1111 CACACCTCTCTGACACCTCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1170

QY 1141 GCGAGCAGGCTTTGATGAGGAGAGATGAGGCTTCTCTCTGCTGAGTGAAGC 1200  
 Db 1171 GCGAGCAGGCTTTGATGAGGAGAGATGAGGCTTCTCTCTGCTGAGTGAAGC 1230  
 QY 1201 AGTCCCTGGGAGTGTGAGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1260  
 Db 1231 AGTCCCTGGGAGTGTGAGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1290  
 QY 1261 GCTTGTGTTTGGCTCTCTTTTGTGTGTCAAGAGCTGCTGCTTCTTGTGAGATGAGA 1320  
 Db 1291 GCTTGTGTTTGGCTCTCTTTTGTGTGTCAAGAGCTGCTGCTTCTTGTGAGATGAGA 1350  
 QY 1321 AGCAGACAGAT 1380  
 Db 1351 AGCAGACAGAT 1410  
 QY 1381 ACTGAGGCTTAGAGGCTGAGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAGAT 1440  
 Db 1411 ACTGAGGCTTAGAGGCTGAGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAGAT 1470  
 QY 1441 GAGGCGGTAAGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1500  
 Db 1471 GAGGCGGTAAGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1530  
 QY 1501 TTTTAAATTAATTAATTAAT 1522  
 Db 1531 TTTTAAATTAATTAATTAAT 1552

RESULT 6  
 US-09-873-367C-516  
 / Sequence 516, Application US/09873367C  
 / Publication No. US20030165839A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Young, Paul  
 / APPLICANT: Soppet, Daniel  
 / APPLICANT: Endress, Gregory  
 / APPLICANT: Augustus, Meena  
 / APPLICANT: Ebner, Reinhard  
 / APPLICANT: Carter, Kenneth  
 / TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using  
 / FILE REFERENCE: 689290-64  
 / CURRENT APPLICATION NUMBER: US/09/873/367C  
 / PRIOR FILING DATE: 2003-04-29  
 / PRIOR APPLICATION NUMBER: U.S. 60/236,891  
 / PRIOR FILING DATE: 2000-09-29  
 / PRIOR APPLICATION NUMBER: U.S. 60/236,842  
 / PRIOR FILING DATE: 2000-09-29  
 / PRIOR APPLICATION NUMBER: U.S. 60/244,867  
 / PRIOR FILING DATE: 2000-11-01  
 / PRIOR APPLICATION NUMBER: U.S. 60/245,084  
 / PRIOR FILING DATE: 2000-11-01  
 / NUMBER OF SEQ ID NOS: 1067  
 / SOFTWARE: PatentIn version 3.0  
 / SEQ ID NO 516  
 / LENGTH: 1552  
 / TYPE: DNA  
 / ORGANISM: Homo sapiens  
 US-09-873-367C-516

Query Match 100.0%; Score 1522; DB 10; Length 1552;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTACAGGAT 60  
 Db 31 ACAGTACAGGAT 90  
 QY 61 CTTGCTCCAGGCTCACTGTGCACTGTGCTGTCACTGTGCTTCTGATGCTGTCCAT 120  
 Db 91 CTTGCTCCAGGCTCACTGTGCACTGTGCTGTCACTGTGCTTCTGATGCTGTCCAT 150



151 |CCCCAGAGTTGCCCCGGATGACAGAGGATTTCCCTTGGAGAGGCTCTTCTGGGAAA| 210  
181 |GATACACCACTGGGCGGAGAGATCTGCCAGAGAGAGATTTCCACCCAGAGAGGAT| 240  
211 |GATACACCACTGGGCGGAGAGATCTGCCAGAGAGAGATTTCCACCCAGAGAGGAT| 270  
241 |CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGAGATTTACTT| 300  
271 |CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGAGATTTACTT| 330  
301 |GAAGTTAGCTTAATAG| 360  
331 |GAAGTTAGCTTAATAG| 390  
361 |GAGGCTCTGAGAGATCTCAAGAACCCAGAGATATAGCCACAGAGAGAGAGAGAG| 420  
391 |GAGGCTCTGAGAGATCTCAAGAACCCAGAGATATAGCCACAGAGAGAGAGAGAG| 450  
421 |GACCAAGATCATTTGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 480  
451 |GACCAAGATCATTTGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 510  
481 |GCGGCGCCTTCCAGTCCCGGATGATATCCGCGCCAGCTCGCCTTCTGCGCGCC| 540  
511 |GCGGCGCCTTCCAGTCCCGGATGATATCCGCGCCAGCTCGCCTTCTGCGCGCC| 570  
541 |CTGGCGCCTTCCAGTCCCGGATGATATCCGCGCCAGCTCGCCTTCTGCGCGCC| 600  
571 |CTGGCGCCTTCCAGTCCCGGATGATATCCGCGCCAGCTCGCCTTCTGCGCGCC| 630  
601 |AACATGGCCACAGAGTGTGACATGACCTGCTCTGAGGCTAGAGATGAGCTCTGGT| 660  
631 |AACATGGCCACAGAGTGTGACATGACCTGCTCTGAGGCTAGAGATGAGCTCTGGT| 690  
661 |GAGCGGAGATACCGGCGCTCTGAGCTGATCTGAGCTGAGGCGCTGAGGCTGAGG| 720  
691 |GAGCGGAGATACCGGCGCTCTGAGCTGATCTGAGCTGAGGCGCTGAGGCTGAGG| 750  
721 |TCGAGGACATCTGTGAGAGGCGACGTTTCCCTGCGGAGATCCAGCTGAGTCACTC| 780  
751 |TCGAGGACATCTGTGAGAGGCGACGTTTCCCTGCGGAGATCCAGCTGAGTCACTC| 810  
781 |ACCGCTTTGCGAGAGTGTGACAGAGGCTTGGGCGCGCGGAGAGGCTTGGCGCTG| 840  
811 |ACCGCTTTGCGAGAGTGTGACAGAGGCTTGGGCGCGCGGAGAGGCTTGGCGCTG| 870  
841 |GCTTTCTGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 900  
871 |GCTTTCTGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 930  
901 |GAGAGATCTGCTGAGAGAGGCTCAGAGATCTAGAGTCCAGAGTCAATATCTGACTC| 960  
931 |GAGAGATCTGCTGAGAGAGGCTCAGAGATCTAGAGTCCAGAGTCAATATCTGACTC| 990  
961 |CTGGCTCTGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 1020  
991 |CTGGCTCTGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 1050  
1021 |GCCAGAGGATGATCTGAGAGTGTGTTTAAACAGAGAGAGAGAGAGAGAGAGAGAG| 1080  
1051 |GCCAGAGGATGATCTGAGAGTGTGTTTAAACAGAGAGAGAGAGAGAGAGAGAGAG| 1110  
1081 |CACACCTCTCTGAGAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 1140  
1111 |CACACCTCTCTGAGAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 1170  
1141 |GCGAGGAGGCTTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 1200  
1171 |GCGAGGAGGCTTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 1230  
1201 |AGTCTGGGCTGCTGAGAGGCTCAGAGTGAATTTCTGGCTGGTGTGATCACTGTT| 1260

1231 |AGTCTGGGCTGCTGAGAGGCTCAGAGTGAATTTCTGGCTGGTGTGATCACTGTT| 1290  
1261 |GCCCTGATTTTGGCTCTCTTTTGTCTGACAGCGTGGCTTCTGTGACAGATGAGA| 1320  
1291 |GCCCTGATTTTGGCTCTCTTTTGTCTGACAGCGTGGCTTCTGTGACAGATGAGA| 1350  
1321 |AGGACGACAG| 1380  
1351 |AGGACGACAG| 1410  
1381 |ACTGAGGCTTAAGAGGCTGAGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAGAG| 1440  
1411 |ACTGAGGCTTAAGAGGCTGAGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAGAG| 1470  
1441 |GAGCGGCTTAAGAGGCTGAGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAGAG| 1500  
1471 |GAGCGGCTTAAGAGGCTGAGATCTTGAAGATGTGAGAGAGAGAGAGAGAGAGAG| 1530  
1501 |TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT| 1522  
1531 |TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT| 1552

RESULT 8  
US-10-465-572-9  
Sequence 9, Application US/10465572  
Publication No. US20030207840A1  
GENERAL INFORMATION:  
APPLICANT: Riggs, Gregory  
TITLE OF INVENTION: GENES INDUCED BY HYPOXIA  
FILE REFERENCE: 000250.00012  
CURRENT FILING DATE: 2003-06-20  
PRIORITY FILING DATE: US/10/201,642  
PRIORITY FILING DATE: 2002-07-24  
PRIORITY FILING DATE: 60/307,600  
PRIORITY FILING DATE: 2001-07-26  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 1552  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-465-572-9

Query Match 100.0%; Score 1522; DB 15; Length 1552;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 |ACAGTACGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG| 60  
31 |ACAGTACGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG| 90  
61 |CTGCTCAGAGGCTCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 120  
91 |CTGCTCAGAGGCTCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 150  
121 |CCCCAGAGGTTGCGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 180  
151 |CCCCAGAGGTTGCGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG| 210  
181 |GATGACCACTGGGCGGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT| 240  
211 |GATGACCACTGGGCGGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT| 270  
241 |CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGAGATTTACTT| 300  
271 |CCACCCGAGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGAGATTTACTT| 330  
301 |GAAGTTAGCTTAATCAG| 360  
331 |GAAGTTAGCTTAATCAG| 390



QY 361 GAGGCTCTGAGATCTCAAGAACCCAGATATATGCCCACAGAGCAAAAGAGGGAT 420  
DB 391 GAGGCTCTGAGATCTCTCAAGAACCCAGATATATGCCCACAGAGCAAAAGAGGGAT 450  
QY 421 GACCAAGATCATTTGGCGCTATGAGAGGCAACCGCCCTGGCCCCGGGTGTCCCAAGCTGC 480  
DB 451 GACCAAGATCATTTGGCGCTATGAGAGGCAACCGCCCTGGCCCCGGGTGTCCCAAGCTGC 510  
QY 481 GCGGCGCGCTGCAATCCCGCGGTGATATCCGCCCCAGCTCGCGCTTCTGCGCGGC 540  
DB 511 GCGGCGCGCTTCAAGTCCCGGTGATATCCGCCCCAGCTCGCGCTTCTGCGCGGC 570  
QY 541 CTGCGCGCCCCGAACTCTGGGCTTCCAGACTCCCGCGCTCCCAAGACTGCGCTGCG 600  
DB 571 CTGCGCGCCCCGAACTCTGGGCTTCCAGACTCCCGCGCTCCCAAGACTGCGCTGCG 630  
QY 601 AACATGCGCAAGTGTGCACTGACCCCGCTCTGGGCTAGAGATGCTCTGGGCTGCC 660  
DB 631 AACATGCGCAAGTGTGCACTGACCCCGCTCTGGGCTAGAGATGCTCTGGGCTGCC 690  
QY 661 GCGCGGAGTACCGGCTCTGAGTGTGCACTGAGCTGGGCGCTGAGTGTGCGGCG 720  
DB 691 GCGCGGAGTACCGGCTCTGAGTGTGCACTGAGCTGGGCGCTGAGTGTGCGGCG 750  
QY 721 TCGAGACACTGTGAAAGGCAACCGTTTCCCTGCGCGAGATCCAGCTGCTCACTGAGC 780  
DB 751 TCGAGACACTGTGAAAGGCAACCGTTTCCCTGCGCGAGATCCAGCTGCTCACTGAGC 810  
QY 781 ACCGCTTTGCCAGATTTGACAGAGCTTGGGCGCCCGGAGGCGCTGCGTGTGGCC 840  
DB 811 ACCGCTTTGCCAGATTTGACAGAGCTTGGGCGCCCGGAGGCGCTGCGTGTGGCC 870  
QY 841 GCGTTTCTGAGAGAGGCGCGGAAAGAAACAATGCTATGAGCAAGTGTGCTGCGCTG 900  
DB 871 GCGTTTCTGAGAGAGGCGCGGAAAGAAACAATGCTATGAGCAAGTGTGCTGCGCTG 930  
QY 901 GAAGAATCGCTAGAGAGGCTCAGAGACTCAGAGCTCCAGAGCTGAGATATCTGCACTC 960  
DB 931 GAAGAATCGCTAGAGAGGCTCAGAGACTCAGAGCTCCAGAGCTGAGATATCTGCACTC 990  
QY 961 CTGCGCTCTGACTTCAAGCGCTATCTTCCATATGAGGGGTCTCTGACTACACCGCTGT 1020  
DB 991 CTGCGCTCTGACTTCAAGCGCTATCTTCCATATGAGGGGTCTCTGACTACACCGCTGT 1050  
QY 1021 GCCCAGGGTGTATCTGAGACTGTGTTAACTGAGAGTGTGCTGAGTGTGCTGAGTGT 1080  
DB 1051 GCCCAGGGTGTATCTGAGACTGTGTTAACTGAGAGTGTGCTGAGTGTGCTGAGTGT 1110  
QY 1081 CACACCTCTCTGACACCTCTGAGAGGAGCTGTGACTCTGAGCTGAGCTGAACTTCCGA 1140  
DB 1111 CACACCTCTCTGACACCTCTGAGAGGAGCTGTGACTCTGAGCTGAGCTGAACTTCCGA 1170  
QY 1141 GCGAGCGAGCTTGAATGGGCGAGATTTAGAGGCTCTTCCCTGCTGAGTGTGAGCAGC 1200  
DB 1171 GCGAGCGAGCTTGAATGGGCGAGATTTAGAGGCTCTTCCCTGCTGAGTGTGAGCAGC 1230  
QY 1201 AGTCTCGAGGCTGTGAGCCAGTCCAGCTGAATTCCTGCTGCTGCTGCTGCTGCTA 1260  
DB 1231 AGTCTCGAGGCTGTGAGCCAGTCCAGCTGAATTCCTGCTGCTGCTGCTGCTGCTA 1290  
QY 1261 GCGCTGCTTTTGGGCGCTTTTGGTGTACACAGAGCTGAGCTTCTTGTGAGATGAGA 1320  
DB 1291 GCGCTGCTTTTGGGCGCTTTTGGTGTACACAGAGCTGAGCTTCTTGTGAGATGAGA 1350  
QY 1321 AGGAGGACAGAGAGGAGGACCAAGAGGAGTGTGAGTACCGCCACAGAGAGTACCGAG 1380  
DB 1351 AGGAGGACAGAGAGGAGGACCAAGAGGAGTGTGAGTACCGCCACAGAGAGTACCGAG 1410  
QY 1381 ACTGAGACCTTGAAGGCTGTGATCTTTGAGAGATGAGAGAGCCAGAGAGGATCTGAGAGG 1440  
DB 1411 ACTGAGACCTTGAAGGCTGTGATCTTTGAGAGATGAGAGAGCCAGAGAGGATCTGAGAGG 1470

QY 1441 GAGCGGCTAACTGCTCTGCTGCTGCTGCTATATGCTCACTTCTTTTAACTGCCAAGAAAT 1500  
DB 1471 GAGCGGCTAACTGCTCTGCTGCTGCTGCTATATGCTCACTTCTTTTAACTGCCAAGAAAT 1530  
QY 1501 TTTTAAATATAATTTTAAAT 1522  
DB 1531 TTTTAAATATAATTTTAAAT 1552

RESULT 9  
US-10-295-027-305  
Sequence 305, Application US/10295027  
Publication No. US20030232350A1  
GENERAL INFORMATION:  
APPLICANT: Afari, Daniel  
APPLICANT: Aziz, Natasha  
APPLICANT: Ginsberg, Wendy M.  
APPLICANT: Gish, Kurt C.  
APPLICANT: Glynn, Richard  
APPLICANT: Hevezi, Peter A.  
APPLICANT: Mack, David H.  
APPLICANT: Murray, Richard  
APPLICANT: Watson, Susan R.  
APPLICANT: Eos Biotechnology, Inc.  
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
FILE REFERENCE: 018501-012500US  
CURRENT FILING DATE: US/10/295, 027  
PRIOR FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US 09/663, 733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350, 666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335, 394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332, 464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334, 393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340, 376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347, 211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347, 349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355, 250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356, 714  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 305  
LENGTH: 1552  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-295-027-305

Query Match 100.0%; Score 1522; DB 15; Length 1552;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTCAAGCGATGCTCCCTGTGCGCCAGCCCTGCTCCCTGTGATCCGAGCC 60  
DB 31 ACAGTCAAGCGATGCTCCCTGTGCGCCAGCCCTGCTCCCTGTGATCCGAGCC 90  
QY 61 CCGTCTCAAGCGCTCACTGTGTGAACTGCTGCTGTCACTGCTGCTTTGATGCGCTCAT 120  
DB 91 CCGTCTCAAGCGCTCACTGTGTGAACTGCTGCTGTCACTGCTGCTTTGATGCGCTCAT 150  
QY 121 CCCAGAGGTTGCCCGAGATGAGAGAGATTCCTCCCTTGGAGAGAGGCTCTTCTGGGAA 180  
DB 151 CCCAGAGGTTGCCCGAGATGAGAGAGATTCCTCCCTTGGAGAGAGGCTCTTCTGGGAA 210







Db	181	GAACCACTGGGGAGAGAGATCTGCCAGTGAAGAGATTCACTCCAGAGAGAGATCC	240					
Qy	244	CCCGGAGAGAGATCTA	CCTGAGAGAGATCTA	CCTGAGAGAGATCTA	CTGAA	303		
Db	241	CCCGGAGAGAGATCTA	CTCCTGAGAGAGAGATCTA	CTGAGAGAGAGATCTA	CTGAA	300		
Qy	304	GTTAAGCCTAATCA	GAAAGAGAGAGGCTCCCTGAAGTTAGAGATCTA	CTACTGTTAG	363			
Db	301	GTTAAGCTTAATCA	GAGAGAGAGGCTCCCTGAAGTTAGAGATCTA	CTACTGTTAG	360			
Qy	364	GCTCTGAGATCTC	CAAGAA	CCCAGAA	TATGCCA	CAAGGACAAAGAGGGATGAC	423	
Db	361	GCTCTGAGATCTC	CAAGAA	CCCAGAA	TATGCCA	CAAGGACAAAGAGGGATGAC	420	
Qy	424	CAGAGCATTTGGGGTAT	TAGAGGGGAC	CCCGCGCTGG	CCCGGGTGGCC	CAGCTTGACG	483	
Db	421	CAGAGCATTTGGGGTAT	TAGAGGGGAC	CCCGCTTGAC	CCCGGGTGGCC	CAGCTTGACG	480	
Qy	484	GGCGCTTCCAGT	CCCCGAGTAT	TCGCCCCCAGCTGG	CGGCTTTCGCCGCGCTG	543		
Db	481	GGCGCTTCCAGT	CCCCGAGTAT	TCGCCCCCAGCTGG	CGGCTTTCGCCGCGCTG	540		
Qy	544	CGCCCCCTGGAATCT	CTGAGGCTTCCAGT	CCCCCGCTCC	CAGAACTGGCGCTTGAC	603		
Db	541	CGCCCCCTGGAATCT	CTGAGGCTTCCAGT	CCCCCGCTCC	CAGAACTGGCGCTTGAC	600		
Qy	604	AATGSCCAGATGTGCA	ACTGACCTCGCTCTG	GGGTAGAGATGCTTGAGT	CCGGG	663		
Db	601	AATGSCCAGATGTGCA	ACTGACCTCGCTCTG	GGGTAGAGATGCTTGAGT	CCGGG	660		
Qy	664	CGGAGATACCGGGCTCT	GACTGATCTGCA	CTGGGGGGCTG	CAGATGCTCCGGGCTG	723		
Db	661	CGGAGATACCGGGCTCT	GACTGATCTGCA	CTGGGGGGCTG	CAGATGCTCCGGGCTG	720		
Qy	724	GAGCACTAGTGAAGG	CCACCGTTCCCTG	CCGAGATCCACGTGTTCA	CTCAGAC	783		
Db	721	GAGCACTAGTGAAGG	CCACCGTTCCCTG	CCGAGATCCACGTGTTCA	CTCAGAC	780		
Qy	784	GCTTTGCCAGACTTGA	CGAGGCTTGGGGG	CCCCGGGAGCCCTGGCGGTTGG	CGGCG	843		
Db	781	GCTTTGCCAGACTTGA	CGAGGCTTGGGGG	CCCCGGGAGCCCTGGCGGTTGG	CGGCG	840		
Qy	844	TTTCTGAGAGAGG	CCCGAGAAACAGT	GGCTTAGAGAGATTTGCTG	CTGCTTGGAA	903		
Db	841	TTTCTGAGAGAGG	CCCGAGAAACAGT	GGCTTAGAGAGATTTGCTG	CTGCTTGGAA	900		
Qy	904	GAAATGCTGAGAGAG	GCTCAGAGCTAGAT	CCCCAGAA	TTGACATATCTGCACCTCG	963		
Db	901	GAAATGCTGAGAGAG	GCTCAGAGCTAGAT	CCCCAGAA	TTGACATATCTGCACCTCG	960		
Qy	964	CCCTGACTTCAGCCG	CTACTTCCAA	TATAGGGG	GTCTGACTAC	CGGCGCTTGCG	1023	
Db	961	CCCTGACTTCAGCCG	CTACTTCCAA	TATAGGGG	GTCTGACTAC	CGGCGCTTGCG	1020	
Qy	1024	CAGGGTGTCA	TTGAGCTGTGTTA	CCAGACAGTGA	TGCTGATGCTTAA	GACAGCTCCAC	1083	
Db	1021	CAGGGTGTCA	TTGAGCTGTGTTA	CCAGACAGTGA	TGCTGATGCTTAA	GACAGCTCCAC	1080	
Qy	1084	ACCTCTCTGAC	CCCTGTGGGAG	CTGATGCTCTG	GGCTACAGCTGAA	CTTCCGAGG	1143	
Db	1081	ACCTCTCTGAC	CCCTGTGGGAG	CTGATGCTCTG	GGCTACAGCTGAA	CTTCCGAGG	1140	
Qy	1144	ACGAGGCTTTGA	TGGGAGATGA	TGAGGCTTC	CTCCGTGGAGTGA	CAGAGAT	1203	
Db	1141	ACGAGGCTTTGA	TGGGAGATGA	TGAGGCTTC	CTCCGTGGAGTGA	CAGAGAT	1200	
Qy	1204	CTTGAGGCTCTAG	CAGCAGTCA	GCTGAATCTG	CTGGCTGCTGGTGA	CATCTAGC	1263	
Db	1201	CTTGAGGCTCTAG	CAGCAGTCA	GCTGAATCTG	CTGGCTGCTGGTGA	CATCTAGC	1260	
Qy	1264	CTGAGTTTGGCTCT	TTTGTGTCA	CACAGCTCGG	GTCTGTTGA	GAGTGAAG	1322	

Db	1261	CTGGTTTTGGCCTCCTTTTTGGCTGACACAGGTCGGCTCTTGTCAGATGAGAAG	1350
Qy	1324	CAGCACAGAGGGGAACTCAAAGGGGGTGTAGCTACCGCCACAGAGGTAGCCGAGCT	1383
Db	1321	CAGCACAAGAGGGGAACCAAGGGGGTGTAGCTACCGCCACAGAGGTAGCCGAGCT	1380
Qy	1384	GGAGCTTAAGGCTGATCTTGAGAAATGAGAAAGCAGCAGAGGCATCTGAGGGGGA	1443
Db	1381	GGAGCTTAAGGCTGATCTTGAGAAATGAGAAAGCAGCAGAGGCATCTGAGGGGGA	1440
Qy	1444	GCCGTAACCTGCTGCTGCTCTGCTATTATGCACTTCCTTTAACTGCCAAGAAATTTT	1503
Db	1441	GCCGTAACCTGCTGCTGCTCTGCTATTATGCCACTTCCTTTAACTGCCAAGAAATTTT	1500
Qy	1504	TAAATTAATATTTATAAT	1522
Db	1501	TAAATTAATATTTATAAT	1519

```

US-09-783-708-2
/ Sequence 2, Application US/09783708
/ Patent No. US20020058041A1
/ GENERAL INFORMATION:
/ APPLICANT: The Regents of the University of California
/ TITLE OF INVENTION: KIDNEY-SPECIFIC TUMOR VACCINE DIRECTED AGAINST KIDNEY TUMOR
/ FILE REFERENCE: 306T-897420US
/ CURRENT APPLICATION NUMBER: US/09/783,708
/ CURRENT FILING DATE: 2001-02-13
/ PRIOR APPLICATION NUMBER: 60/182,429
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: 60/182,636
/ PRIOR FILING DATE: 2000-02-15
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: Patentln version 3.0
/ SEQ ID NO 2
/ LENGTH: 1833
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Nucleic acid encoding G250-GW-CSF fusion protein with His tag
US-09-783-708-2

Query Match          90.7%; Score 1379.8; DB 9; Length 1833;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1381; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      16  GCTCCCTGTGACCCCAAGCCCTGGCTCCCTGTGTGATCCCGGCCCTGCTCAGGCTC 75
DB      439  GCTCCCTGTGACCCCAAGCCCTGGCTCCCTGTGTGATCCCGGCCCTGCTCAGGCTC 498

QY      76  ACTGTGCACACTGTGCTGTACTGTCTGCTTGTGATGCTGTTCATCCCCAGAGTTCCCC 135
DB      499  ACTGTGCACACTGTGCTGTACTGTCTGCTTGTGATGCTGTTCATCCCCAGAGTTCCCC 558

QY      136  CGGATGCAAGAGATATCCCTTGGAGAGAGGCTCTTGGGGAAAGATGATCCCACTGAGC 195
DB      559  CGGATGCAAGAGATATCCCTTGGAGAGAGGCTCTTGGGGAAAGATGATCCCACTGAGC 618

QY      196  GAGAGAGATCTGCCAGTGAAGAGATTCAACCACAGAGAGAGATTCACCCGAGAGAG 255
DB      619  GAGAGAGATCTGCCAGTGAAGAGATTCAACCACAGAGAGAGATTCACCCGAGAGAG 678

QY      256  GATCTACTGGAAGAGAGATCTTAACCTGGAAGAGAGATCTTAACCTGAAGTTAAGCTAAA 315
DB      679  GATCTACTGGAAGAGAGATCTTAACCTGGAAGAGAGATCTTAACCTGAAGTTAAGCTAAA 738

QY      316  TCAGAGAGAGAGGCTCCCTGAAGTAGAGAGATTACTACTGTGAGGCTCCTGAGAT 375
DB      739  TCAGAGAGAGAGGCTCCCTGAAGTAGAGAGATTACTACTGTGAGGCTCCTGAGAT 798

QY      376  CCTCAGAACCCCAAGATTATGCCACAGGACAAAGAGGAGTGAACCAAGTCATTGC 435

```

Db 799 CCTCAAGAACCCAGATATATGCCCAAGAGGAAAGGAGATGACCAAGATCATTTGG 858  
 Qy 436 CGCTATGAGAGGACCCGCTGAGCCCGGAGTGTCCCAAGCTGCGCGGCGCTTCCAG 495  
 Db 859 CGCTATGAGAGGACCCGCTGAGCCCGGAGTGTCCCAAGCTGCGCGGCGCTTCCAG 918  
 Qy 496 TCCCGGAGTATTCGCGCCCAAGCTGCGCTTGTCCCGGAGCTTGTGCGCCCTGGA 555  
 Db 919 TCCCGGAGTATTCGCGCCCAAGCTGCGCTTGTCCCGGAGCTTGTGCGCCCTGGA 978  
 Qy 556 CTCTGGGCTTCAAGCTCCGCGCTCCAGAACTGCGCTGCGCAACAATGAGCCAGT 615  
 Db 979 CTCTGGGCTTCAAGCTCCGCGCTCCAGAACTGCGCTGCGCAACAATGAGCCAGT 1038  
 Qy 616 GTGCACTGACCTGCTCTGAGGCTAGAGATGCTGAGTCTGAGTCTGCGGCGGAGTACCG 675  
 Db 1039 GTGCACTGACCTGCTCTGAGGCTAGAGATGCTGAGTCTGAGTCTGCGGCGGAGTACCG 1098  
 Qy 676 GCTGTGAGCTGATCTGCACTGCGGCGGCTGCGAGTCTGCGGCGGCTGCGAGCACTGTG 735  
 Db 1099 GCTGTGAGCTGATCTGCACTGCGGCGGCTGCGAGTCTGCGGCGGCTGCGAGCACTGTG 1158  
 Qy 736 GAGGCGACCGTTTCCCTGCGAGATCCAGTGTTCACCTGACACCGCTTGTCCAGA 795  
 Db 1159 GAGGCGACCGTTTCCCTGCGAGATCCAGTGTTCACCTGACACCGCTTGTCCAGA 1218  
 Qy 796 GTTACGAGGCTTGGGCGCGCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 855  
 Db 1219 GTTACGAGGCTTGGGCGCGCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1278  
 Qy 856 GCGCGGAGAAACAGTGTGCTATGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 915  
 Db 1279 GCGCGGAGAAACAGTGTGCTATGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1338  
 Qy 916 GAAGGCTGAGACTCAAGTCCAGAGCTGGAATATCTGCACTCTGCGCTTGTGACTTC 975  
 Db 1339 GAAGGCTGAGACTCAAGTCCAGAGCTGGAATATCTGCACTCTGCGCTTGTGACTTC 1398  
 Qy 976 AGCGGCTACTCCAAATGAGAGGCTCTGACTACACCGCGCTGTGCCAGAGGTTCATC 1035  
 Db 1399 AGCGGCTACTCCAAATGAGAGGCTCTGACTACACCGCGCTGTGCCAGAGGTTCATC 1458  
 Qy 1036 TGAAGTGTATTAACAAGAGAGTGTGCTGAGAGCTGCAACCTCTGTGAC 1095  
 Db 1459 TGAAGTGTATTAACAAGAGAGTGTGCTGAGAGCTGCAACCTCTGTGAC 1518  
 Qy 1096 ACCCTGAGGAGACCTGAGTCTGCGGCTACAGTGAATTCGAGAGCAAGAGCTTGTG 1155  
 Db 1519 ACCCTGAGGAGACCTGAGTCTGCGGCTACAGTGAATTCGAGAGCAAGAGCTTGTG 1578  
 Qy 1156 AATGGGAGATGATGAGGCTCTCTTCTGAGTGAAGAGAGTCTGCGGCTGCT 1215  
 Db 1579 AATGGGAGATGATGAGGCTCTCTTCTGAGTGAAGAGAGTCTGCGGCTGCT 1638  
 Qy 1216 GAGGAGTCCAGAGTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1275  
 Db 1639 GAGGAGTCCAGAGTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1698  
 Qy 1276 CTCTCTTTTGTCTACAGAGCTGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1335  
 Db 1699 CTCTCTTTTGTCTACAGAGCTGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1758  
 Qy 1336 GGAACCAAGAGGCTGTAGTACCTGCGCGAGAGAGTACCGAGCTGAGGCTTGAAG 1395  
 Db 1759 GGAACCAAGAGGCTGTAGTACCTGCGCGAGAGAGTACCGAGCTGAGGCTTGAAG 1818  
 Qy 1396 CTG 1398  
 Db 1819 GTG 1821

RESULT 13  
 US-09-772-719-5

; Sequence 5, Application US/09772719  
 ; Patent No. US20020137910A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zavada, Jan  
 ; APPLICANT: Pastorekova, Silvia  
 ; APPLICANT: Pastorek, Jaromir  
 ; TITLE OF INVENTION: MN Gene and Protein  
 ; NUMBER OF SEQUENCES: 86  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Leona L. Lauder  
 ; STREET: 369 Pine Street  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/772,719  
 ; FILING DATE: 30-JAN-2001  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/485,049  
 ; FILING DATE: 07-JUN-1995  
 ; APPLICATION NUMBER: US 08/260,190  
 ; FILING DATE: 15-JUN-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lauder, Leona L.  
 ; REGISTRATION NUMBER: 30,863  
 ; REFERENCE/DOCKET NUMBER: D-0021.3E  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-981-2034  
 ; TELEFAX: 415-981-0332  
 ; INFORMATION FOR SEQ. ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 10898 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULAR TYPE: DNA (genomic)  
 ; HYPOTHEICAL: NO  
 ; ANTI-SENSE: NO  
 ; US-09-772-719-5  
 Query Match 27.2%; Score 414.4; DB 9; Length 10898;  
 Best Local Similarity 99.8%; Pred. No. 9.3e-112;  
 Matches 415; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 ACHGTACGCCGATGAGTCTCCCTGTGCCCCAGCCCTGCTCCCTGTGATCCCGGCC 60  
 Db 3537 ACHGTACGCCGATGAGTCTCCCTGTGCCCCAGCCCTGCTCCCTGTGATCCCGGCC 3596  
 Qy 61 CCGCTCCAGGCTCACTGTGCACTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 Db 3597 CCGCTCCAGGCTCACTGTGCACTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3656  
 Qy 121 CCCGAGGTTGCCCCGAGATGCAAGAGATTTCCCTTGGAGAGAGGCTCTTCTGAGGA 180  
 Db 3657 CCCGAGGTTGCCCCGAGATGCAAGAGATTTCCCTTGGAGAGAGGCTCTTCTGAGGA 3716  
 Qy 181 GATGACCACTGGGCGAGAGATCTGCCAGTGAAGAGATTCCACGAGAGAGAGAT 240  
 Db 3717 GATGACCACTGGGCGAGAGATCTGCCAGTGAAGAGATTCCACGAGAGAGAGAT 3776  
 Qy 241 CCACCGGAGAGAGATCTTACCTTGAGAGAGAGATCTTCTGAGAGAGAGATCTTACT 300  
 Db 3777 CCACCGGAGAGAGATCTTACCTTGAGAGAGAGATCTTCTGAGAGAGAGATCTTACT 3836  
 Qy 301 GAAGTTAGGCTAAATCAGAGAGAGAGGCTCCCTGAAGTGAAGATCTTACTGTT 360

Db 3637 GAAGTTAAGCTTAATCAGAGAAGAGGCTCCCTGAAGTTAGAGGATCTACTGTT 3696  
QY 361 GAGGCTCTGGAGATCTTCAAGAACCCCAATTAATGCCACAGGAGCAAGAGG 416  
Db 3697 GAGGCTCTGGAGATCTTCAAGAACCCCAATTAATGCCACAGGAGCAAGAGG 3952

## RESULT 14

US-09-967-237-5  
Sequence 5, Application US/09967237  
Publication No. US20030049828A1  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
FILE REFERENCE: D-0021.5B-2  
CURRENT APPLICATION NUMBER: US/09/967,237  
PRIORITY FILING DATE: 2001-09-27  
PRIORITY FILING DATE: 1998-10-23  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 10898  
TYPE: DNA  
ORGANISM: HUMAN  
FEATURE:  
NAME/KEY: gene  
LOCATION: (1)-(10898)  
OTHER INFORMATION: Full-length MN genomic sequence  
NAME/KEY: unsure of base at position 1974  
LOCATION: (1974)  
OTHER INFORMATION: unsure of base at position 1974, which is in the 5' region flanking  
OTHER INFORMATION: transcription initiation site (3507) as determined by RNase probe  
US-09-967-237-5

Query Match 27.2%; Score 414.4; DB 10; Length 10898;

Best Local Similarity 99.8%; Pred. No. 9,3e-112;

Matches 415; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAGTCAGCCGATGAGTCCCTGTCGCCAGCCCTGAGCTCCCTGTGATCCCGGCC 60  
Db 3537 ACAGTCAGCCGATGAGTCCCTGTCGCCAGCCCTGAGCTCCCTGTGATCCCGGCC 3596  
QY 61 CTTGCTCCAGGCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
Db 3597 CTTGCTCCAGGCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3656  
QY 121 CCCGAGAGGTTGCCCGCGATGAGAGGATTTCCCTTGGAGAGGCTCTTCTGGGGA 180  
Db 3557 CCCGAGAGGTTGCCCGCGATGAGAGGATTTCCCTTGGAGAGGCTCTTCTGGGGA 3716  
QY 181 GATGACCCACTGGGCGAGAGGATTTGCCCAATTAATTAATTAATTAATTAATTAAT 240  
Db 3117 GATGACCCACTGGGCGAGAGGATTTGCCCAATTAATTAATTAATTAATTAATTAAT 3776  
QY 241 CCACCGGAGAGGATTTACTGAGAGGATTTACTGAGAGGATTTACTGAGAGGATTTACT 300  
Db 3777 CCACCGGAGAGGATTTACTGAGAGGATTTACTGAGAGGATTTACTGAGAGGATTTACT 3896  
QY 301 GAAGTTAAGCTTAATCAGAGAAGAGGCTCCCTGAAGTTAGAGGATTTACTGAGGTT 360  
Db 3897 GAAGTTAAGCTTAATCAGAGAAGAGGCTCCCTGAAGTTAGAGGATTTACTGAGGTT 3896  
QY 361 GAGGCTCTGGAGATCTTCAAGAACCCCAATTAATTAATTAATTAATTAATTAAT 416  
Db 3897 GAGGCTCTGGAGATCTTCAAGAACCCCAATTAATTAATTAATTAATTAATTAAT 3952

RESULT 15  
US-09-772-719-28  
Sequence 28, Application US/09772719

Parent No. US20020137910A1

GENERAL INFORMATION:

APPLICANT: Zavada, Jan

APPLICANT: Pastorekova, Silvia

APPLICANT: Pastorek, Jaromir

TITLE OF INVENTION: MN Gene and Protein

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSER: Leona L. Lauder

STREET: 369 Pine Street

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/772,719

FILING DATE: 30-JAN-2001

CLASSIFICATION:

Prior Application Data:

APPLICATION NUMBER: US 08/485,049

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: US 08/260,190

FILING DATE: 15-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.

REGISTRATION NUMBER: 30,863

REFERENCE/DOCKET NUMBER: D-0021.3B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-981-0332

TELEFAX: 415-981-2034

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 415 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

DESCRIPTION: 1st MN exon

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-09-772-719-28

Query Match 27.2%; Score 413.4; DB 9; Length 415;

Best Local Similarity 99.8%; Pred. No. 4,8e-112;

Matches 414; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAGTCAGCCGATGAGTCCCTGTCGCCAGCCCTGAGCTCCCTGTGATCCCGGCC 60  
Db 1 ACAGTCAGCCGATGAGTCCCTGTCGCCAGCCCTGAGCTCCCTGTGATCCCGGCC 60  
QY 61 CTTGCTCCAGGCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
Db 61 CTTGCTCCAGGCTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
QY 121 CCCGAGAGGTTGCCCGAGATGAGAGGATTTCCCTTGGAGAGGCTCTTCTGGGGA 180  
Db 121 CCCGAGAGGTTGCCCGAGATGAGAGGATTTCCCTTGGAGAGGCTCTTCTGGGGA 180  
QY 181 GATGACCCACTGGGCGAGAGGATTTGCCCAATTAATTAATTAATTAATTAATTAAT 240  
Db 181 GATGACCCACTGGGCGAGAGGATTTGCCCAATTAATTAATTAATTAATTAATTAAT 240  
QY 241 CCACCGGAGAGGATTTACTGAGAGGATTTACTGAGAGGATTTACTGAGAGGATTTACT 300  
Db 241 CCACCGGAGAGGATTTACTGAGAGGATTTACTGAGAGGATTTACTGAGAGGATTTACT 300  
QY 301 GAAGTTAAGCTTAATCAGAGAAGAGGCTCCCTGAAGTTAGAGGATTTACTGAGGTT 360

Db 301 GAAGTTAGCCTAAATCAGAAGAGAGGCTCCCTGAAGTTAGAGATCTACCTACTGTT 360  
Qy 361 GAGGCTCTGTGAGATCTCTCAAGAAACCCAGATATATGCCCCACAGGACAAAGAG 415  
Db 361 GAGGCTCTGTGAGATCTCTCAAGAAACCCAGATATATGCCCCACAGGACAAAGAG 415

Search completed: March 7, 2004, 06:33:57  
Job time : 422 secs





http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paradey Avenue Genoscope sequence ID : CS0D1085C09QPI.  
Location/Qualifiers

## FEATURES

Source

1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1085F18"  
/tissue\_type="PLACENTA"  
/clone\_lib="Homo sapiens PLACENTA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

## ORIGIN

Query Match 69.2%; Score 1053.8; DB 9; Length 1201;  
Best Local Similarity 96.2%; Pred. No. 3,7e-224;  
Matches 1086; Conservative 21; Mismatches 19; Indels 3; Gaps 3;

1 ACAGTCAGCCGATGCTCCCTGTCGCCAGCCCTGCTCCCTCTGTTGATCCGAGC 60  
75 AAGATCAGCCGATGCTCCCTGTCGCCAGCCCTGCTCCCTCTGTTGATCCGAGC 133  
61 CCTGCTCAGCCGATGCTCCCTGTCGCCAGCCCTGCTCCCTCTGTTGATCCGAGC 120  
134 CCTGCTCAGCCGATGCTCCCTGTCGCCAGCCCTGCTCCCTCTGTTGATCCGAGC 193  
121 CCCGAGAGGTTGCCCGGATGAGAGAGATCCCGCTTGGAGAGAGCTCTTGGAGAA 180  
194 CCCGAGAGGTTGCCCGGATGAGAGAGATCCCGCTTGGAGAGAGCTCTTGGAGAA 253  
181 GATGACCACTGGGAGAGAGATGTCGCCAGTGAAGAGATTCACCCAGAGAGAGAT 240  
254 GATGACCACTGGGAGAGAGATGTCGCCAGTGAAGAGATTCACCCAGAGAGAGAT 313  
241 CCACCCGAGAGAGAGATGTCGCCAGTGAAGAGATTCACCCAGAGAGAGAT 300  
314 CCACCCGAGAGAGAGATGTCGCCAGTGAAGAGATTCACCCAGAGAGAGAT 373  
301 GAAGTTAAGCTTAATCAGAGAGAGAGATGTCGCCAGTGAAGAGATTCACCTGTT 360  
374 GAAGTTAAGCTTAATCAGAGAGAGAGATGTCGCCAGTGAAGAGATTCACCTGTT 433  
361 GAGCTCTCTGAGAGATGTCGCCAGTGAAGAGATTCACCCAGAGAGAGAT 420  
434 GAGCTCTCTGAGAGATGTCGCCAGTGAAGAGATTCACCCAGAGAGAGAT 493  
421 GACCAAGTCAATGAGAGATGTCGCCAGTGAAGAGATTCACCCAGAGAGAT 480  
494 GACCAAGTCAATGAGAGATGTCGCCAGTGAAGAGATTCACCCAGAGAGAT 553  
481 GCGGAGCGCTTCAAGTCCCGGATGAGATTCGCCAGCTGCGCGCTTGTGCCGCGC 540  
554 GCGGAGCGCTTCAAGTCCCGGATGAGATTCGCCAGCTGCGCGCTTGTGCCGCGC 613  
541 CTGCGCCCTTGAACTCTCGGCTTCCAGCTCCCGCGCTCCAGAACTGCGCTGCGC 600  
614 CTGCGCCCTTGAACTCTCGGCTTCCAGCTCCCGCGCTCCAGAACTGCGCTGCGC 673  
601 AACATGCGCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 660  
674 AACATGCGCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 733  
661 GGGCGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 720  
734 GGGCGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 793  
721 TCGAGAGCACTGTGAAGGCAACGTTTCCCTGCGCGAGATTCACGTGTTCACCTGAGC 780  
794 TCGAGAGCACTGTGAAGGCAACGTTTCCCTGCGCGAGATTCACGTGTTCACCTGAGC 853

781 ACCGCTTTGCCAGAGTTCAGAGGCTTTGGAGGCGCCCGGAGAGGCTTGGAGC 840  
854 ACCGCTTTGCCAGAGTTCAGAGGCTTTGGAGGCGCCCGGAGAGGCTTGGAGC 913  
841 GCGTTTCTGAGAGAGGCGCCCGGAGAGAGAGTGGCTTATGACATTTGCTGCTTG 900  
914 GCGTTTCTGAGAGAGGCGCCCGGAGAGAGAGTGGCTTATGACATTTGCTGCTTG 973  
901 GAGAAATCCGCTGAGAGAGGCTTCAGAGTCAAGTCCAGAGC-TGACATATCTGACT 959  
974 GAGAAATCCGCTGAGAGAGGCTTCAGAGTCAAGTCCAGAGTCAAGTCAAGTCAAGT 1033  
960 CTGCGCTCTGACTTCAGAGGCGCTTCCAAATATGAGGGGCTTCAGCTACAGCGCCG 1019  
1034 CTGCGCTCTGACTTCAGAGGCGCTTCCAAATATGAGGGGCTTCAGCTACAGCGCCG 1093  
1020 TGCCAGAGTGTATCTGAGACTGTGTTTACAGAGAGTGAAGTGAAGTGAAGTGAAGT 1079  
1094 TGCCAGAGTGTATCTGAGACTGTGTTTACAGAGAGTGAAGTGAAGTGAAGTGAAGT 1153  
1080 CCACACCTCTCTGACACCTCTGAGGAGCTGTGATCTCGGCTTACAG 1128  
1154 CCACACCTCTCTGACACCTCTGAGGAGCTGTGATCTCGGCTTACAG 1201

RESULT 2  
AL554705 1201 bp mRNA linear EST 31-MAY-2003  
LOCUS  
DEFINITION  
AL554705 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0D1085F18 5-PRIME, mRNA sequence.

ACCESSION  
AL554705  
VERSION  
AL554705.2 GI:31276515  
KEYWORDS  
EST.

SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
Li W.B., Gruber C., Jessee J., and Polayes J.  
Full-length cDNA libraries and normalization  
Unpublished (2001)

JOURNAL  
On Feb 15, 2001 this sequence version replaced gi:12895743.  
Contact: Genoscope  
Genoscope - Centre National de Sequenage  
BP 191 91006 Evry cedex - France

http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D1085D09QPI&cluster=5300.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paradey Avenue Genoscope sequence ID : CS0D1085C09QPI.

FEATURES  
Location/Qualifiers

Source

1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1085F18"  
/tissue\_type="PLACENTA"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 60.5%; Score 920.8; DB 9; Length 1201;  
Best Local Similarity 96.4%; Pred. No. 1.8e-194;  
Matches 991; Conservative 13; Mismatches 15; Indels 9; Gaps 6;

1 ACAGTCAGCCGATGCTCCCTGTCGCCAGCCCTGCTCCCTCTGTTGATCCGAGC 60

```

Db      |||
99  ACAGTAGCGAGTGGCTCCCTGTGCCCCAGCCCTGGCTCCCTGTGATCCCGGCC 158
Qy      |||
61  CCCTGCTCCAGGCTCTCATGTGAACTGCTGCTGTACTGCTGCTCTGTAAGCCGTGCAT 120
Db      |||
159  CCGTCTCCAGGCTCTCATGTGAACTGCTGCTGTACTGCTGCTCTGTAAGCCGTGCAT 218
Qy      |||
121  CCCCAGAGGTTGCCCCGATGCAAGAGATTCCTCCCTTGGAGAGAGGCTCTTGGGAA 180
Db      |||
219  CCCCAGAGGTTGCCCCGATGCAAGAGATTCCTCCCTTGGAGAGAGGCTCTTGGGAA 278
Qy      |||
181  GATGACCCCACTGGGCGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT 338
Db      |||
279  GATGACCCCACTGGGCGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGAT 338
Qy      |||
241  CCACCCGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 300
Db      |||
339  CCACCCGAGAGAGATCTACTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 398
Qy      |||
301  GAAGTTAAGCTTAATCAAGAAAGAGAGGCTCCCTGAACTTGAAGATCTACTACTGTT 360
Db      |||
399  GAAGTTAAGCTTAATCAAGAAAGAGAGGCTCCCTGAACTTGAAGATCTACTACTGTT 458
Qy      |||
361  GAGGCTCTGAGAGATCTCAAGAAAGAGAGATTAATGCCCCAGAGAGAGAGAGAT 420
Db      |||
459  GAGGCTCTGAGAGATCTCAAGAAAGAGAGATTAATGCCCCAGAGAGAGAGAGAT 518
Qy      |||
421  GACCAGAGTCAATTGGCGCTATGAGAGCGACCCGCGCTGGCCCGGCTGTCCCAAGCTTC 480
Db      |||
519  GACCAGAGTCAATTGGCGCTATGAGAGCGACCCGCGCTGGCCCGGCTGTCCCAAGCTTC 578
Qy      |||
481  GCGGCGCGCTTCCAGTCCCGGCTGATATCCGCGCCCGAGCTGCGCGCTTCCGCGGCC 540
Db      |||
579  GCGGCGCGCTTCCAGTCCCGGCTGATATCCGCGCCCGAGCTGCGCGCTTCCGCGGCC 638
Qy      |||
541  CTGCGCCCTCTGGAATCTCTGGGCTTCCAGCTCCCGCGCTCCAGAACTGCGCTGCGC 600
Db      |||
639  CTGCGCCCTCTGGAATCTCTGGGCTTCCAGCTCCCGCGCTCCAGAACTGCGCTGCGC 698
Qy      |||
601  AACATGCGCACTGTGCACTGACCTGCTCCCTGGGCTGAGAGATGGCTTGGGCTCC 660
Db      |||
699  AACATGCGCACTGTGCACTGACCTGCTCCCTGGGCTGAGAGATGGCTTGGGCTCC 758
Qy      |||
661  GCGCGGAGTACCAGGCTCTGCACTGCACTGCACT- GCGGCGCTGCAAGTCTGCGG 719
Db      |||
759  GCGGCGGAGTACCAGGCTCTGCACTGCACTGCACTGCACTGCGGCGGCTGCGG 818
Qy      |||
720  CTGCGAGCACTGTGGAAGGCCACGCTTCCCTGCGGAGATCCAGTGTGCTCACTCAG 779
Db      |||
819  CTGCGAGCACTGTGGAAGGCCACGCTTCCCTGCGGAGATCCAGTGTGCTCACTCAG 878
Qy      |||
780  CACGCGCTTCCAGAGTGTGCAAGAGGCTTGGGCGCGCCGCGAGAGCTGCGCGTGTGC 839
Db      |||
879  CACGCGCTTCCAGAGTGTGCAAGAGGCTTGGGCGCGCCGCGAGAGCTGCGCGTGTGC 938
Qy      |||
840  CGCCTTCTGAGAGAGGCGCCGGAAGAAACAAGTGCCTATGAGCACTGTGCTGCTT 899
Db      |||
939  CGCCTTCTGAGAGAGGCG- CCGGAGAGAAACAAGTGCCTATGAGCACTGTGCTGCTT 997
Qy      |||
900  GGAAGAAATGCTGAGAGAGGCTCAGAGACTCAGAGTCCAGAGCTGAGCACTATGCACT 959
Db      |||
998  GGAAGAAATGCTGAGAGAG- GCTCAGAGACTCAGAGTCCAGAGCTGAGCACTATGCACT 1056
Qy      |||
960  CTTGCGCTCTGACTTACCGCTTCAATATGAGGAGTCTCTGACTTACACCGGCTTG 1019
Db      |||
1057  C-TGGCTCTGACTTACG-CGCTATTTCCAMAMAGAGGCTCTT---KATWMAACCGGCTTG 1110
Qy      |||
1020  TGGCCAGG 1027
Db      |||
1111  TGGCAGGG 1118

```

```

BX423970
LOCUS      BX423970      1013 bp      mRNA      linear      EST 15-MAY-2003
DEFINITION BX423970 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CS0DA003YB12 5-PRIME, mRNA sequence.
ACCESSION  BX423970
VERSION     BX423970.1 GI:30766328
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  Li W.B., Gruber C., Jesse J., and Polayes D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by life technologies, a division of
            invitrogen. This sequence belongs to sequence cluster 5300.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CSIDA001ZF09QPL&cluster=5300.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSIDA001ZF09QPL.

FEATURES
    source
        1..1013
            location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSDA003YB12"
                /tissue_type="NEUROBLASTOMA"
                /clone_id="Homo sapiens NEUROBLASTOMA"
                /note="Vector: pCMVSPORT.5; 1st strand cDNA was primed
                with a NotI-oligo(dT) primer. Five prime end enriched,
                double-strand cDNA was digested with Not I and cloned into
                the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                Library was not normalized."

ORIGIN
Query Match      60.1%; Score 915; DB 13; Length 1013;
Best Local Similarity 94.8%; Pred. No. 3.2e-193; Indels 0; Gaps 0;
Matches 906; Conservative 33; Mismatches 19;

2  CAGTCAGCCGATGCTCCCTGTGCCCCAGCCCTGGCTCCCTGTGATCCCGGCC 61
Db      |||
56  CCGGAGTCCGATAGCTCCCTGTGCCCCAGCCCTGGCTCCCTGTGATCCCGGCC 115
Qy      |||
62  CTGCTCAAGGCTCATGTGCACTGCTGCTGCACTGCTGCTGTAAGCTGTCATC 121
Db      |||
116  CTGCTCAAGGCTCATGTGCACTGCTGCTGCACTGCTGCTGTAAGCTGTCATC 175
Qy      |||
122  CCCAGAGTGGCCCGATGCAAGAGATTCCTCCCTTGGAGAGAGCTTCTGGGAAG 181
Db      |||
176  CCCAGAGTGGCCCGATGCAAGAGATTCCTCCCTTGGAGAGAGCTTCTGGGAAG 235
Qy      |||
182  ATGACCACTGGGCGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGATC 241
Db      |||
236  ATGACCACTGGGCGAGAGATCTGCCAGTGAAGAGATTCACCCAGAGAGAGATC 295
Qy      |||
242  CACCCGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGAGATCTACCTG 301
Db      |||
296  CACCCGAGAGAGATCTACCTGAGAGAGATCTACCTGAGAGAGATCTACCTG 355
Qy      |||
302  AAGTTAAGCTTAATCAAGAAAGAGGCTCCCTGAACTTGAAGATCTACTACTGTTG 361
Db      |||
356  AAGTTAAGCTTAATCAAGAAAGAGGCTCCCTGAACTTGAAGATCTACTACTGTTG 415
Qy      |||
362  AGGCTCTGAGAGATCCCAAGAACCCAGAAATATCCCAAGAGAGAGAGAGATG 421
Db      |||
416  AGGCTCTGAGAGATCCCAAGAACCCAGAAATATCCCAAGAGAGAGAGAGATG 475

```

```

QY 422 ACCAGAGTATTGGCGCTATGAGAGCGACCCGCTGCGCCCGGGTGTCCCAAGCTTGG 481
DB 476 ACCAGAGTATTGGCGCGATGAGAGCGACCCGCTGCGCCCGGGTGTACCAAGCTTGG 545
QY 482 CGGCGCGCTTCCAGATCCCGGATGATATCCGCGCCAGCTGCGCGCTTGTGCGCGGCC 541
DB 536 CGGCGCGCTTCCAGATCCCGGATGATATCCGCGCCAGCTGCGCGCTTGTGCGCGGCC 595
QY 542 TGGCGCCCTGGAATCTTGGGCTTCAAGCTTCCGCGCTGCGCGCTTGTGCGCGGCC 601
DB 596 TGGCGCCCTGGAATCTTGGGCTTCAAGCTTCCGCGCTGCGCGCTTGTGCGCGGCC 655
QY 602 ACAATGGCCACAGTGTGCAACTGACCTGCTGCGGGGTGAGATGAGCTTGTGCGCGGCC 661
DB 656 ACAATGGCCACAGTGTGCAACTGACCTGCTGCGGGGTGAGATGAGCTTGTGCGCGGCC 715
QY 662 GGGCGGAGTACCGGGGCTTGCAGAGCTGATCTGCACTGGGGGCTGCGAGTGTGCGGGCT 721
DB 716 GGGCGGAGTACCGGGGCTTGCAGAGCTGATCTGCACTGGGGGCTGCGAGTGTGCGGGCT 775
QY 722 CGAGGACACTGTGAGAGGCGACCGCTTCCGCGAGATCCAGTGTGCTGACCTGACGA 781
DB 776 CGAGGACACTGTGAGAGGCGACCGCTTCCGCGAGATCCAGTGTGCTGACCTGACGA 835
QY 782 CGGCTTTGGCAGAGTGTGAGAGGCGCTTGGGGCGCGCGAGGCTTGGCGCTTGTGGCG 841
DB 836 CGGCTTTGGCAGAGTGTGAGAGGCGCTTGGGGCGCGCGAGGCTTGGCGCTTGTGGCG 895
QY 842 CTTTCTGAGAGAGGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
DB 896 CTTTCTGAGAGAGGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955
QY 902 AAGAAATCCGCTGAGAGAGCTGAGACTCCAGACTCCAGACTGAGACTATCTGACT 959
DB 956 AAAAAAMGCTGAGAGAGCTCAGAAATTTAGATTTAGATTTAGATTTAGATTTAGAT 1013

RESULT 4
AL558378 1072 bp mRNA linear EST 31-MAY-2003
LOCUS AL558378 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0D005YK14 5-PRIME, mRNA sequence.
ACCESSION AL558378
VERSION AL558378.2 GI:31280177
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 1072)
L1,W.B., Gruber,C., Jessee,J. and Polayee,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12902842.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 101 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5300.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D005BF07QPLcluster=5300.f. Contact :
Peng Liang Email : filiang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope, sequence ID : CS0D005BF07QPL.
FEATURES
Source
1..1072
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D005YK14"

```

```

/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/clone_id="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 55.4%; Score 903.4; DB 9; Length 1072;
Best Local Similarity 99.1%; Pred. No. 1,3e-190;
Matches 956; Conservative 4; Mismatches 0; Indels 5; Gaps 5;
QY 1 ACAGTCAGCCGATGCTGCTCCCTGTGCGCCGAGCCCTGCTGCTGTTGATCCGAGC 60
DB 94 ACAGTCAGCCGATGCTGCTCCCTGTGCGCCGAGCCCTGCTGCTGTTGATCCGAGC 153
QY 61 CTTGCTCAGGCTTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
DB 154 CTTGCTCAGGCTTCACTGTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 213
QY 121 CCCGAGAGTTGCCCGGATGACAGAGATTTCCCTTGGAGAGAGCTTTTGGAGAA 180
DB 214 CCCGAGAGTTGCCCGGATGACAGAGATTTCCCTTGGAGAGAGCTTTTGGAGAA 273
QY 181 GATGACCCACTGGGCGAGAGAGATCTCCAGTGAAGAGATTTCACAGAGAGAGAT 240
DB 274 GATGACCCACTGGGCGAGAGATCTCCAGTGAAGAGATTTCACAGAGAGAGAT 333
QY 241 CCACCCGAGAGAGAGATCTTACGAGAGAGAGATTTACCTGAGAGAGAGATTTACT 300
DB 334 CCACCCGAGAGAGAGATCTTACGAGAGAGAGATTTACTGAGAGAGAGATTTACT 393
QY 301 GAAGTTAAGCTTAATCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 360
DB 394 GAAGTTAAGCTTAATCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 453
QY 361 GAGGCTCTGAGAGATCTTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 420
DB 454 GAGGCTCTGAGAGATCTTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 513
QY 421 GACCAAGATCAATGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480
DB 514 GACCAAGATCAATGGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 573
QY 481 GCGGCGCGCTTCAAGATCTTCAAGAGATTTCCGCGCGAGAGAGAGAGAGAGAGAT 540
DB 574 GCGGCGCGCTTCAAGATCTTCAAGAGATTTCCGCGCGAGAGAGAGAGAGAGAGAT 633
QY 541 CTGGCGCCCTGGAATCTTGGGCTTCCAGCTCCCGCGCTGCCAGAACTGCGCTGCG 600
DB 634 CTGGCGCCCTGGAATCTTGGGCTTCCAGCTCCCGCGCTGCCAGAACTGCGCTGCG 693
QY 601 AACAAATGGCCACAGTGTGCAACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 694 AACAAATGGCCACAGTGTGCAACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
QY 661 GGGCGGAGTACCGGGGCTCTGCAAGCTGATCTGCACTGGGGGCTGCAAGTGTGCGG 720
DB 754 GGGCGGAGTACCGGGGCTCTGCAAGCTGATCTGCACTGGGGGCTGCAAGTGTGCGG 813
QY 721 TGGAGACACTGTGGAAGAGCCAGCTTCCCTGCGAGATCCAGTGTGCTGCTGCTGCT 780
DB 814 TGGAGACACTGTGGAAGAGCCAGCTTCCCTGCGAGATCCAGTGTGCTGCTGCTGCT 873
QY 781 ACCGCTTTGCGAGAGTGAAGAGAGCTTGGGGGCGCGCGAGAGAGAGAGAGAGAG 840
DB 874 ACCGCTTTGCGAGAGTGAAGAGAGCTTGGGGGCGCGCGAGAGAGAGAGAGAGAG 922
QY 841 GCGCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 900
DB 933 GCGCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 991

```





QY 513 CCCCAGCTGCGCGCTTCTGCGCGGCGCTGCGCGGCGCTGCGGACTCTGAGGCTTCCAGCT 572  
 Db 985 CCGCCCCCGCTGCGCGCTTCTGCGCGGCGCGGCGCGGCGCTGAG--CTCTGAGTTCGMBT 928  
 QY 573 CCGCGCGCTTCCAGAACTGCGCGCTGCGCAACTGCGCAAGTGTGCACTGAGCCCTGCC 632  
 Db 927 CCGCGCGCTTCCAGAACTGCG--CTGCGCTATATAGGCAAGTGTGCTATYRACCCCTCC 869  
 QY 633 TCTGAGGCTAGAGATGCTCTGAGTCCGCGGCGGAGTACCAGGCTCTGAGCTGCACT 692  
 Db 868 TCTGAGGCTATAGATGCTCTGAGTCCGCGGCGGAGTACCAGGCTCTGAGCTGCACT 809  
 QY 693 GCACTGCGGCGCTGCGAGTGTGCGGCGCTGCGAGCACTGTGAGAGGCGCTTCC 752  
 Db 808 GAACTGCGGCGCTGCGAGTGTGCGGCGCTGCGAGCACTGTGAGAGGCGCTTCC 749  
 QY 753 TCGCGAGTCCAGTGTGCTCACTGAGCAAGCGCTTCCAGAGTTGACAGGCGCTTCC 812  
 Db 748 TCGCGAGTCCAGTGTGCTCACTGAGCAAGCGCTTCCAGAGTTGACAGGCGCTTCC 689  
 QY 813 GCGCGCGGAGGCGCTGCGCGCTTCTGCGCGGCGCTTCTGAGAGGCGCGGAGAAACAG 872  
 Db 688 GCGCGCGGAGGCGCTGCGCGCTTCTGCGCGGCGCTTCTGAGAGGCGCGGAGAAACAG 629  
 QY 873 TCGCTATGAGCACTGTGCTGCTGCTTGGAGAAATCGCTGAGAGGCTCAGAGCTCA 932  
 Db 628 TCGCTATGAGCACTGTGCTGCTGCTTGGAGAAATCGCTGAGAGGCTCAGAGCTCA 569  
 QY 933 GGTCCAGAGACTGAGCAATCTGCACTCTGCGCTTCTGAGAGGCGCTTCTGCAATA 992  
 Db 568 GGTCCAGAGACTGAGCAATCTGCACTCTGCGCTTCTGAGAGGCGCTTCTGCAATA 509  
 QY 993 TGAAGGCTCTGCACTACACCGCGCGCTGCGAGGCTGCACTGCACTGCACTGCACT 1052  
 Db 508 TGAAGGCTCTGCACTACACCGCGCGCTGCGAGGCTGCACTGCACTGCACTGCACT 449  
 QY 1053 GACAGTATGCTGAGTGTGAGAGGCTGCACTGCACTGCACTGCACTGCACTGCACT 1112  
 Db 448 GACAGTATGCTGAGTGTGAGAGGCTGCACTGCACTGCACTGCACTGCACTGCACT 389  
 QY 1113 TGACTTGGCTTACAGTGAATCTTCCAGAGGAGGCGCTTGGATGGGCGAGTATGA 1172  
 Db 388 TGACTTGGCTTACAGTGAATCTTCCAGAGGAGGCGCTTGGATGGGCGAGTATGA 329  
 QY 1173 GAGCTCTTCCCTGCTGAGTGAAGAGCAAGTCTGCGGCTGCTGAGGCGAGTCAAGTGA 1232  
 Db 328 GAGCTCTTCCCTGCTGAGTGAAGAGCAAGTCTGCGGCTGCTGAGGCGAGTCAAGTGA 269  
 QY 1233 TTTCTGCTGAGTGTGAGTGAATCTTCCAGAGGAGTGTGAGTGTGCTGCTGCTGCT 1292  
 Db 268 TTTCTGCTGAGTGTGAGTGAATCTTCCAGAGGAGTGTGAGTGTGCTGCTGCTGCT 209  
 QY 1293 CAGGCTGAGTGTGAGTGAAGAGCAAGTCTGCGGCTGCTGAGGCGAGTCAAGTGA 1352  
 Db 208 CAGGCTGAGTGTGAGTGAAGAGCAAGTCTGCGGCTGCTGAGGCGAGTCAAGTGA 149  
 QY 1353 GAGTACCGGCGAGAGTGAAGAGCAAGTCTGAGAGGCTTGAAGTGTGAGTGAATG 1412  
 Db 148 GAGTACCGGCGAGAGTGAAGAGCAAGTCTGAGAGGCTTGAAGTGTGAGTGAATG 89  
 QY 1413 TGAAGAGCGGCGAGAGTGAAGAGGAGGCGGAGTGTGAGTGTGCTGCTGCT 1462  
 Db 88 TGAAGAGCGGCGAGAGTGAAGAGGAGGCGGAGTGTGAGTGTGCTGCTGCT 39

RESULT 8  
 AL577748/c 1031 bp mRNA linear EST 01-JUN-2003  
 LOCUS AL577748 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
 DEFINITION CDNA clone CS0DK007YK10 3-PRIME, mRNA sequence.  
 ACCESSION AL577748  
 VERSION AL577748.2 GI:31316001

KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 1031)  
 AUTHORS Li, W.B., Gruber, C., Jessup, J., and Polayres, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 16, 2001 this sequence version replaced gi:12941176.  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by life technologies, a division of  
 invitrogen. This sequence belongs to sequence cluster 5300.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DK007BF05NP1&cluster=5300.f. Contact :  
 Feng Liang Email : fliang@life.techn.com URL :  
 http://fulllength.invitrogen.com/invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DK007BF05NP1.  
 FEATURES  
 source  
 1..1031  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DK007YK10"  
 /cell\_type="HELA"  
 /cell\_line="HELA"  
 /clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-clig(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 ORIGIN  
 Query Match 52.4%; Score 796.8; DB 9; Length 1031;  
 Best Local Similarity 86.0%; Pred. No. 7.9e-167;  
 Matches 852; Conservative 50; Mismatches 84; Indels 5; Gaps 4;  
 QY 521 TCGCGCGCTTCCGCGCGCGCTGCGCGGCGCTGCGGACTCTGAGCTTCCGCGCGC 580  
 Db 989 TCGCGCGCTTCCGCGCGCGCTTCCGCGCGCGCTGCGGACTCTGAGCTTCCGCGCGC 930  
 QY 581 TCCGAGACTGCGCGCTGCGCAACTGCGCAAGTGTGCACTGAGCTGCTGCGCGC 640  
 Db 929 CYCCGAGVACTCTCTGCAKCAKAGGCGCA--KGTGCAATGTGCTTSCYGGYKGGC 871  
 QY 641 TGAAGTGTGCTGAGGCTTCCGCGCGCGAGTACCGGCTCTGAGCTGAGTCT--GCACTG 699  
 Db 870 KAGAGTGTGCTGAGGCTTCCGCGCGCGAGTACCGGCTCTGAGCTGAGTCT--GCACTG 811  
 QY 700 GGGGCTGAGGCTGCTGCGCGCGCTGAGAGCACTGTGAGAGGCAAGCTTCTGCGCGAG 759  
 Db 810 GAGGCTGAGGCTGCTGCGCGCGCTGAGAGCACTGTGAGAGGCAAGCTTCTGCGCGAG 751  
 QY 760 ATCCAGTGTGCTGAGGCTTCCGCGCGCGCTTCCGCGCGAGTGTGAGAGGCTTCCGCGCGAG 819  
 Db 750 ATCCAGTGTGCTGAGGCTTCCGCGCGCGCTTCCGCGCGAGTGTGAGAGGCTTCCGCGCGAG 692  
 QY 820 GAGGCTGAGGCTTCCGCGCGCGCTTCCGCGCGAGTGTGAGAGGCTTCCGCGCGAG 879  
 Db 691 GAGGCTGAGGCTTCCGCGCGCGCTTCCGCGCGAGTGTGAGAGGCTTCCGCGCGAG 632  
 QY 880 GAGGCTGAGGCTTCCGCGCGCGCTTCCGCGCGAGTGTGAGAGGCTTCCGCGCGAG 939  
 Db 631 GAGGCTGAGGCTTCCGCGCGCGCTTCCGCGCGAGTGTGAGAGGCTTCCGCGCGAG 572  
 QY 940 GAGGCTGAGGCTTCCGCGCGCGCTTCCGCGCGAGTGTGAGAGGCTTCCGCGCGAG 999  
 Db 571 GAGGCTGAGGCTTCCGCGCGCGCTTCCGCGCGAGTGTGAGAGGCTTCCGCGCGAG 512



```

QY 1000 TCTGTACATACACGCCCTGTGACGAGGAGTGTCTGATGTGTTTAAACAGACAGTG 1059
Db 511 TCTGTACATACACGCCCTGTGACGAGGAGTGTCTGATGTGTTTAAACAGACAGTG 452
QY 1060 ATGCTGAGTGTACAGAGCTCCACACCTCTCTGACACCTCTGTGGGAGCTGTGATCTC 1119
Db 451 ATGCTGAGTGTACAGAGCTCCACACCTCTCTGACACCTCTGTGGGAGCTGTGATCTC 392
QY 1120 CGGCTACAGCTGAACCTCCGAGCGACGAGCTTTGATGGGAGATGATTGAGGCGCTCC 1179
Db 391 CGGCTACAGCTGAACCTCCGAGCGACGAGCTTTGATGGGAGATGATTGAGGCGCTCC 332
QY 1180 TTCCCTGTGAGTGTGACAGCAGCTCTGAGGCTGTGAGCCAGTCCAGCTGATTCCTGC 1239
Db 331 TTCCCTGTGAGTGTGACAGCAGCTCTGAGGCTGTGAGCCAGTCCAGCTGATTCCTGC 272
QY 1240 CTGGCTGTGAGTGTGACATCTCTAGCCTGGGTTTGGCCCTTTTGTGCTGACAGGCTC 1299
Db 271 CTGGCTGTGAGTGTGACATCTCTAGCCTGGGTTTGGCCCTTTTGTGCTGACAGGCTC 212
QY 1300 GCGTTCCTTGTGAGATGAGAAAGAGACACAGAGGAGAACCAAGGGGCTGTGAGCTAC 1359
Db 211 GCGTTCCTTGTGAGATGAGAAAGAGACACAGAGGAGAACCAAGGGGCTGTGAGCTAC 152
QY 1360 CGGCCAGAGAGGTAGCCGAGAGCTGAGGCTGAGCTGATCTTGTGAGAAATGAGAG 1419
Db 151 CGGCCAGAGAGGTAGCCGAGAGCTGAGGCTGAGCTGATCTTGTGAGAAATGAGAG 92
QY 1420 CCAGCCAGAGGATCTGAGGAGGAGCCGATACCTGTCTGTCTCTCATTAAGCA--C 1477
Db 91 CCAGCCAGAGGATCTGAGGAGGAGCCGATACCTGTCTGTCTCTCATTAAGCA--C 32
QY 1478 TTCTTTTAACTGCGCAAGAAATTTTAAAA 1508
Db 31 TCTATTTAATGTCWNTNNNTTTTMMW 1

```

```

RESULT 9
AL580216/c 1017 bp mRNA linear EST 01-JUN-2003
LOCUS AL580216 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0D005YK14 3-PRIME, mRNA sequence.
ACCESSION AL580216
VERSION AL580216.2 GI:31318495
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1017)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization.
JOURNAL Unpublished (2001)
COMMENT On Feb 16, 2001 this sequence version replaced gi:12946023.
Contact: Genoscope - Centre National de Sequencage
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5300.f for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D005BF07NP1&cluster=5300.f. Contact:
Peng Liang Email: filiang@lifetech.com URL:
http://fulllength.invitrogen.com/invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CS0D005BF07NP1.
Location/Qualifiers
1..1017
/organism="Homo sapiens"
/mo_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D005YK14"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"

```

```

/orig_line="JURKAT"
/clone_id="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/orig="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and cloned into the NotI and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

```

Query Match 51.9%; Score 789.4; DB 9; Length 1017;
Best local similarity 87.9%; Pred. No. 3.5e-165;
Matches 832; Conservative 44; Mismatches 66; Indels 4; Gaps 3;

```

```

QY 545 GCGCCCTGGAACCTCTGAGGCTCCAGTCCCGCGCTCCAGAACTCCGCTCGGACACA 604
Db 943 GCGCCCTGGAACCTCTGAGGCTCCAGTCCCGCGCTCCAGAACTCCGCTCGGACACA 886
QY 605 ATGGCCAGTGTGCACTGACCTGCTCTCTGAGTGAAGTGTGCTGTGGGTCCGGG 663
Db 885 ATGGCCAGTGTGCACTGACCTGCTCTCTGAGTGAAGTGTGCTGTGGGTCCGGG 826
QY 664 CGGAGATACCGGGCTCTGAGCTGATCTGACCTGAGGAGGCTGACAGTGTCCGGGCTCG 723
Db 825 CGGAGATACCGGGCTCTGAGCTGATCTGACCTGAGGAGGCTGACAGTGTCCGGGCTCG 766
QY 724 GAGCACTGTGAAAGGCCACCGTTTCTGCGGAGATCCAGTGTTCACCTCAGCACC 783
Db 765 GAGCACTGTGAAAGGCCACCGTTTCTGCGGAGATCCAGTGTTCACCTCAGCACC 706
QY 784 GCGTTTGGCAAGTGTGACGAGGCTTGGGAGGCGCGGAGAGGCTGAGGCTGTGGGCGCG 843
Db 705 VCGAAACCCAGATGTGAGAGGCTTGGGAGGCGCGGAGAGGCTGAGGCTGTGGGCGCG 646
QY 844 TTTCTGAGAGGAGGCGCGGAGAGAGAGAGTGTCTGAGAGTGTCTGTCTGAGAA 903
Db 645 TTTCTGAGAGGAGGCGCGGAGAGAGAGAGTGTCTGAGAGTGTCTGTCTGAGAA 586
QY 904 GAAATCGCTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGT 963
Db 585 GAAATCGCTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGT 526
QY 964 CCCTCTGACTCAGCGGCTTACTTCCAAATATAGAGGAGTCTGTACTACAGCCCTGTGCG 1023
Db 525 CCCTCTGACTCAGCGGCTTACTTCCAAATATAGAGGAGTCTGTACTACAGCCCTGTGCG 466
QY 1024 CAGGCTGTGATGTGAGTGTGATTTAAACAGAGTGTGAGTGTGAGTGTGAGTGTGAG 1083
Db 465 CAGGCTGTGATGTGAGTGTGATTTAAACAGAGTGTGAGTGTGAGTGTGAGTGTGAG 406
QY 1084 ACCCTCTGAGACACCTGTGAGGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1143
Db 405 ACCCTCTGAGACACCTGTGAGGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 346
QY 1144 ACAGAGCTTTGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1203
Db 345 ACAGAGCTTTGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 286
QY 1204 CCTCGGAGCTGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1263
Db 285 CCTCGGAGCTGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 226
QY 1264 CTGCTTTTGGCTCTCTTTTGTGCTGACAGGCTGTGCTGCTGTGAGTGTGAGTGTGAG 1323
Db 225 CTGCTTTTGGCTCTCTTTTGTGCTGACAGGCTGTGCTGCTGTGAGTGTGAGTGTGAG 166
QY 1324 CAGCAGAGAGGAGAACCAAGGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1383
Db 165 CAGCAGAGAGGAGAACCAAGGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 107
QY 1384 GAGAGCTAGAGGCTGTGATCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1443
Db 106 GAGAGCTAGAGGCTGTGATCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 47

```







QY 909 CGCTGAGGAAGGCTCAGAGACTCAGGTCCAGAGACTGACATATCTGACTCTGCTCC 968  
 Db 629 CGCTGAGGAAGGCTCAGAGACTCAGGTCCAGAGACTGACATATCTGACTCTGCTCC 570  
 QY 969 TGACTTACGCGCTACTCTCCAAATATAGAGGGGTCTGATATAACCGCCCTGTGCCAGG 1028  
 Db 569 TGAATTGAGCGCTACTCTCCAAATATAGAGGGGTCTGATATAACCGCCCTGTGCCAGG 510  
 QY 1029 TGATCTGAGTGTGTGTTTAAACAGACAGTGAATGCTGAATGCTGAAGAGCTCCACACCT 1088  
 Db 509 TGATCTGAGTGTGTGTTTAAACAGACAGTGAATGCTGAATGCTGAAGAGCTCCACACCT 450  
 QY 1089 CTCTGACACCCCTGTGGGACCTGTGATCTCTGGCTTACAGCTGAACCTTCCAGACGCA 1148  
 Db 449 CTCTGACACCCCTGTGGGACCTGTGATCTCTGGCTTACAGCTGAACCTTCCAGACGCA 390  
 QY 1149 GCCTTGAATGGGAGAGTGAATGAGGCTCTTCCCTGCTGAGTGAAGAGAGAGTCTCG 1208  
 Db 389 GCCTTGAATGGGAGAGTGAATGAGGCTCTTCCCTGCTGAGTGAAGAGAGAGTCTCG 330  
 QY 1209 GGTGCTGAGCCAGTCCAGCTGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1268  
 Db 329 GGTGCTGAGCCAGTCCAGCTGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 270  
 QY 1269 TTTTGGCTCTCTTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1328  
 Db 269 TTTTGGCTCTCTTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 210  
 QY 1329 CAGAGGGGGAACCAAGGGGGGTGTGAGCTACCGCCAGCAGAGAGTGAAGCTGAGAGC 1388  
 Db 209 CAGAGGGGGAACCAAGGGGGGTGTGAGCTACCGCCAGCAGAGAGTGAAGCTGAGAGC 150  
 QY 1389 CTGAGGCTGAGTCTTGAAGAAATGAGAAAGCAGCAGAGAGTGAAGGGGAGCCG 1448  
 Db 149 CTGAGGCTGAGTCTTGAAGAAATGAGAAAGCAGCAGAGAGTGAAGGGGAGCCG 90  
 QY 1449 TAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1508  
 Db 89 TAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 30  
 QY 1509 TAAATATTTTAAAT 1522  
 Db 29 TAAATATTTTAAAT 16

RESULT 13  
 A1831707/c 682 bp mRNA linear EST 21-DEC-1999  
 LOCUS DEFINITION wj40h01.x1 NCI CGAP Lul9 Homo sapiens cDNA clone IMAGE:2405329 3'  
 similar to TR:Q16790 Q16790 CARBONIC ANHYDRASE 1, mRNA sequence.  
 A1831707  
 A1831707.1 GI:5452464  
 EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 682)  
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Straube, Ph.D.  
 Email: gspds-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 clone distribution: NCI-CCAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ULNL at:  
 www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 1322 Std Error: 0.00  
 Seq primer: -40UP from Gldco  
 High quality sequence stop: 424.  
 Location/Qualifiers  
 1..682  
 source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2405329"  
 /tissue\_type="squamous cell carcinoma, poorly  
 differentiated (4 pooled tumors, including primary and  
 metastatic)"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NCI-CCAP\_Lul9"  
 /note="Organ: Lung; Vector: pRT73D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from  
 pooled lung tumor tissue, and was then primed with a Not I  
 - oligo(dT) primer. Double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pRT73 vector. Library went through one round of  
 normalization. Library constructed by Bento Soares and M.  
 Fatima Bonaldo."

ORIGIN

Query Match 43.7%; Score 665.8; DB 9; Length 682;  
 Best Local Similarity 98.8%; Pred. No. 1.1e-137;  
 Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 845 TTCTGAGAGAGGCGCCGGAAGAAACAGTGCCTATGACAGTTCCTGCTGCTGAG 904  
 Db 678 TTTTGAAGAAAGGCGGGAAGAAACAGTGCCTATGACAGTTCCTGCTGCTGAG 619  
 QY 905 AATGCTGAGAAAGCTCAAGACTCAGGCTCCAGAGCTGAGATATCTGCACTCTGCG 964  
 Db 618 ACATGCTGAGAAAGCTCAAGACTCAGGCTCCAGAGCTGAGATATCTGCACTCTGCG 559  
 QY 965 CCTGACTTCAAGCGCTACTTCCAAATGAGGGGCTCTGACTACACCGCCCTGTGCC 1024  
 Db 558 CTTTGACTTCAAGCGCTACTTCCAAATGAGGGGCTCTGACTACACCGCCCTGTGCC 499  
 QY 1025 AGGATGCTATCTGACTGTGTTTAAACAGAAAGTGAAGTGAAGAGCTCCACA 1084  
 Db 498 AGGATGCTATCTGACTGTGTTTAAACAGAAAGTGAAGTGAAGAGCTCCACA 439  
 QY 1085 CCTCTCTGACACCTGTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1144  
 Db 438 CCTCTCTGACACCTGTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 379  
 QY 1145 CGAGCTTTGAATGGGAGAGTGAAGGCTCTTCCCTGCTGAGTGAAGAGAGTGC 1204  
 Db 378 CGAGCTTTGAATGGGAGAGTGAAGGCTCTTCCCTGCTGAGTGAAGAGAGTGC 319  
 QY 1205 CTGAGGCTGCTGAGCAGTCCAGTGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1264  
 Db 318 CTGAGGCTGCTGAGCAGTGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 259  
 QY 1265 TGGATTTTGGCTCTCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1324  
 Db 258 TGGATTTTGGCTCTCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 199  
 QY 1325 AGCAGAGAGGGAACCAAGGGGGGTGAGCTACCGCCAGCAGAGAGTGAAGAGAGTGC 1384  
 Db 198 AGCAGAGAGGGAACCAAGGGGGGTGAGCTACCGCCAGCAGAGAGTGAAGAGAGTGC 139  
 QY 1385 GAGCTTGAAGGCTGATCTTGAAGATGTAAGAACCCAGCAGAGAGTGAAGAGAGTGC 1444  
 Db 138 GAGCTTGAAGGCTGATCTTGAAGATGTAAGAACCCAGCAGAGAGTGAAGAGAGTGC 79  
 QY 1445 CGGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1504  
 Db 78 CGGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19

QY 1505 AAATAAATATTATAT 1522  
 DB 18 AAATAAATATTATAT 1

RESULT 14  
 BX383092 1074 bp mRNA linear EST 08-MAY-2003  
 LOCUS BX383092 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
 DEFINITION Homo sapiens cDNA clone CS0DJ005YGI0 5-PRIME, mRNA sequence.  
 BX383092  
 ACCESSION EX383092.1 GI:30449113  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 5300.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DJ005BD05GPI&cluster=5300.f. Contact :  
 Feng Liang Email: fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DJ005BD05GPI.  
 Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DJ005YGI0"  
 /cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
 /cell\_line="JURKAT"  
 /clone\_1fb="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-clisio (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
 Query Match 43.7%; Score 665.4; DB 13; Length 1074;  
 Best Local Similarity 93.1%; Pred. No. 1.7e-137;  
 Matches 734; Conservative 14; Mismatches 6; Indels 34; Gaps 4;

1 ACAGTACGCGATGCTCCCTGTCGCCCAACCCCTGCTCTTGTATCCCGCC 60  
 DB 70 ACAGTACGCGATGCTCCCTGTCGCCCAACCCCTGCTCTTGTATCCCGCC 129

QY 61 CCTGCTCCAGGCTCACTGTGCAACTGCTGCTCACTGCTGCTTGTATCCCTGTCAT 120  
 DB 130 CCTGCTCCAGGCTCACTGTGCAACTGCTGCTGCTCACTGCTGCTTGTATCCCTGTCAT 189

QY 121 CCCGAGAGGTTGCCCGCGAGTGCAGAGAGATCCCTTGGAGAGAGGCTCTTCTGGGAA 180  
 DB 190 CCCGAGAGGTTGCCCGCGAGTGCAGAGAGATCCCTTGGAGAGAGGCTCTTCTGGGAA 249

QY 181 GATGACCCACTGGCGAGAGAGATCTGCCAAGTGAAGAGATTACCCAGAGAGAGAT 240  
 DB 250 GATGACCCACTGGCGAGAGAGATCTGCCAAGTGAAGAGATTACCCAGAGAGAGAT 309

QY 241 CCACCCGAGAGAGAGATCTACCTGAGAGAGAGATCTACCTGAGAGAGAGATCTACT 300  
 DB 310 CCACCCGAGAGAGATCTACCTGAGAGAGAGATCTACTGAGAGAGAGATCTACT 369

QY 301 GAAGTTAAGCTTAATATCAAGAGAGAGGCTCCCTGAAGTTAGAGATCTACTACTGT 360  
 DB 370 GAAGTTAAGCTTAATATCAAGAGAGAGGCTCCCTGAAGTTAGAGATCTACTACTGT 429

QY 361 GAGGCTCTGAGAGCTCTCAAGAACCCCAAGTAATATGCCACAGGAGACAAAGAGGGAT 420  
 DB 430 GAGGCTCTGAGAGCTCTCAAGAACCCCAAGTAATATGCCACAGGAGACAAAGAGGGAT 482

QY 421 GACCAGAGTATTGGGCTATGAGAGGACACCCGCTGAGCCCGGGGTGTGCCCAAGCTGC 480  
 DB 483 -----AGCGAAGCCGCTGTGCTCCCGGATGTCCCAAGCTGC 519

QY 481 GCGGAGCGCTTCCAGTCCCGGCTGATATCCGCGCCACAGCTGCGCGCTTCTGCCGCGC 540  
 DB 520 GCGGAGCGCTTCCAGTCCCGGCTGATATCCGCGCCACAGCTGCGCGCTTCTGCCGCGC 579

QY 541 CTGCGCCCTCTGAACTCTGAGCTTCCAGCTTCCGCGCTCCAGAACTGCGGCTGCGC 600  
 DB 580 CTGCGCCCTCTGAACTCTGAGCTTCCAGCTTCCGCGCTCCAGAACTGCGGCTGCGC 639

QY 601 AACATGAGCAAGGTGCAACTGACACCTGCTCTGAGCTAGAGATGCTGAGTCC 660  
 DB 640 AACATGAGCAAGGTGCAACTGACACCTGCTCTGAGCTAGAGATGCTGAGTCC 699

QY 661 GCGGAGAGTACCGGCTCTGACAGCTGATCTGACCTGAGGCTGACAGTGTGTCGCGC 720  
 DB 700 GCGGAGAGTACCGGCTCTGACAGCTGATCTGACCTGAGGCTGACAGTGTGTCGCGC 759

QY 721 TCGAGACACACTGTGAGAGGCAACCGTTTCCCGGAGATCAGCTGTGACCTCAGC 780  
 DB 760 TC-GGRSMATGTGGAAAGACACS--TTTCTCTCCAGAT-CMGTGTTCACCTCAGC 815

QY 781 ACCGCTT 788  
 DB 816 ACGCTT 823

RESULT 15  
 BU620600/c 691 bp mRNA linear EST 23-SEP-2002  
 LOCUS BU620600 UI-H-FL1-bfx-o-01-0-UI.s1 NCI CGAP FL1 Homo sapiens cDNA clone  
 DEFINITION UI-H-FL1-bfx-o-01-0-UI 3', mRNA sequence.  
 BU620600  
 ACCESSION BU620600.1 GI:23286815  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryotic Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 Tissue Procurement: James Martin  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strauszberg, Ph.D.  
 Email: cga@bs-remail.nih.gov  
 Tissue Procurement: James Martin  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 CDNA Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 1-45, >AT rich\_low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
 Location/Qualifiers  
 1..691  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FL1-bfx-o-01-0-UI"

```
/tissue_type="Cell lines"  
/dev_stage="Adult"  
/lab_host="DH10B (Life Technologies)"  
/clone_lib="NCI CGAP FL1"  
/notes="Organ: Chondrosarcoma; Vector: p773-Pac  
(Pharmacia) with a modified polylinker; Site 1: Ecor I;  
Site 2: Not I; NCI CGAP FL1 is a normalized cDNA library  
derived from a pool of mRNA obtained from 4 cell lines  
from grade III chondrosarcoma tissues. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an Ecor I  
adaptor, digested with Not I, and cloned directionally  
into p773-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
GAGTGGCTG. The cell lines were provided by Dr. James  
Martin from the University of Iowa.  
TAG_TISSUE=human Chondrosarcoma Grade 3 cell line mix  
TAG_LIB=UI-H-FL1  
TAG_SEQ=GAGTGGCTG"
```

## ORIGIN

```
Query Match 43.4%; Score 660.4; DB 13; Length 691;  
Best Local Similarity 99.7%; Pred. No. 1.8e-136;  
Matches 672; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

```
QY 849 GGAGGAGGCGCCGGAAGAAACAGTGCCTATGAGCAGTGTCTCTGCTTGGAGAAAT 908  
DB 691 GGAGGAGGCGCCGGAAGAAACAGTGCCTATGAGCAGTGTCTCTGCTTGGAGAAAT 632  
QY 909 CGCTGAGGAGGCTCAGAGACTAGGTCCAGACTGAGACATATCTGACCTCCGCTC 968  
DB 631 CGCTGAGGAGGCTCAGAGACTAGGTCCAGACTGAGACATATCTGACCTCCGCTC 572  
QY 969 TGAATTGAGCGGCTACTTCCATATATAGGGGTCTCTGACTACCGCCCTGTGCCAGG 1028  
DB 571 TGAATTGAGCGGCTACTTCCATATATAGGGGTCTCTGACTACCGCCCTGTGCCAGG 512  
QY 1029 TGTGATCTGAGCTGTGTTTAAACGACAGTATGATGCTGATGCTTAAAGCAGCTCCACCT 1088  
DB 511 TGTGATCTGAGCTGTGTTTAAACGACAGTATGATGCTGATGCTTAAAGCAGCTCCACCT 452  
QY 1089 CTCTGACACCCCTGAGGAGCCTGCTGACTCTGAGCTACAGCTGAACCTCCAGCAGCGCA 1148  
DB 451 CTCTGACACCCCTGAGGAGCCTGCTGACTCTGAGCTACAGCTGAACCTCCAGCAGCGCA 392  
QY 1149 GCCTTTGAATGGGCGAGTATGAGGCTTCCTTCCCTGCTGAGTGAAGCAGCAGCTCTCG 1208  
DB 391 GCCTTTGAATGGGCGAGTATGAGGCTTCCTTCCCTGCTGAGTGAAGCAGCAGCTCTCG 332  
QY 1209 GAGCTGCTGAGCAGCTCAGCTGAATTCCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1268  
DB 331 GAGCTGCTGAGCAGCTCAGCTGAATTCCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 273  
QY 1269 TTTTGGCCTCCTTTTGTGCTACCAAGCGTGCCTTCTTGTGAGATGAGAGGACACA 1328  
DB 272 TTTTGGCCTCCTTTTGTGCTACCAAGCGTGCCTTCTTGTGAGATGAGAGGACACA 213  
QY 1329 CAGAGGAGGAGCAACAAAGGGGGTGTGAGCTACCGCCGAGAGAGGTAGCCGAGACTGAGAGC 1388  
DB 212 CAGAGGAGGAGCAACAAAGGGGGTGTGAGCTACCGCCGAGAGAGGTAGCCGAGACTGAGAGC 153  
QY 1389 CTAGAGGCTGATCTTGGAGAAATGTGAGAAAGCCAGCAGAGGCAATCTGAGGGGAGCCGG 1448  
DB 152 CTAGAGGCTGATCTTGGAGAAATGTGAGAAAGCCAGCAGAGGCAATCTGAGGGGAGCCGG 93  
QY 1449 TAACTGCTGCTCTGCTCATTTATGCACTTCTTTTAACTGCCAAGAAATTTTAAAA 1508  
DB 92 TAACTGCTGCTCTGCTCATTTATGCACTTCTTTTAACTGCCAAGAAATTTTAAAA 33
```

```
QY 1509 TAAATATTTATAT 1522  
DB 32 TAAATATTTATAT 19
```

```
Search completed: March 7, 2004, 05:23:47  
Job time: 2834 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using SW model

Run on: March 5, 2004, 08:58:24 ; Search time 59 Seconds  
(without alignments)  
2198.124 Million cell updates/sec

Title: US-09-967-237-2

Perfect score: 2424  
Sequence: 1 MAPLCSPMLPILIPAPAPQ.....RRGTGVSYPRAVAETGA 459

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1980s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2424	100.0	459	2	AAR8058 Protein e
2	2424	100.0	459	3	AAV53228 Human MN
3	2424	100.0	459	3	AAAB03005 Human MN
4	2424	100.0	459	5	AAE17175 Human RCC
5	2424	100.0	459	6	ABR58596 Human can
6	2424	100.0	459	6	ABP97744 Arthro aci
7	2424	100.0	459	6	ABU56556 Lung can
8	2424	100.0	459	6	ABU56408 Lung can
9	2419	99.8	459	6	ABR82848 Kidney ca
10	2007	82.8	377	3	AAV53245 Human protei
11	2007	82.8	377	3	AAAB03021 Mature hu
12	1398.5	57.7	429	2	AAAR41746 MN protei
13	1370	56.5	257	3	AAV53241 Novel hum
14	1370	56.5	257	3	AAAB03018 Human MN
15	1370	56.5	257	3	AAAR97235 Muru puta
16	1364	56.3	256	2	AAAR97235 Human PRO
17	562	23.2	337	4	AAAB6209 Protein o
18	562	23.2	337	4	AAU29157 Human PRO
19	562	23.2	337	5	AAAB87562 Human PRO
20	562	23.2	337	5	ABG95887 Human sec
21	562	23.2	337	6	ABU58533 Human PRO
22	562	23.2	337	6	ABU84396 Human sec
23	562	23.2	337	6	ABU84396 Human sec
24	562	23.2	337	6	ABR66270 Human sec
25	562	23.2	337	6	ABR65660 Human sec

26	562	23.2	337	6	ABU99600 Human sec
27	562	23.2	337	6	ABU82839 Human PRO
28	562	23.2	337	6	ABU89960 Novel hum
29	562	23.2	337	6	ABR68209 Human sec
30	562	23.2	337	6	ABU96262 Novel hum
31	562	23.2	337	6	ABU92693 Human sec
32	562	23.2	337	6	ABO08770 Human sec
33	562	23.2	337	6	ABO02822 Human sec
34	562	23.2	337	6	ABR74976 Human sec
35	562	23.2	337	6	ABR94738 Human PRO
36	562	23.2	337	6	ABU85711 Human PRO
37	562	23.2	337	6	ABU98871 Novel hum
38	562	23.2	337	6	ABU98086 Novel hum
39	562	23.2	337	6	ABU91792 Novel hum
40	562	23.2	337	6	ABU89485 Human PRO
41	562	23.2	337	6	ABU86326 Human sec
42	562	23.2	337	6	ABU67539 Human sec
43	562	23.2	337	6	ABU80567 Human PRO
44	562	23.2	337	6	ABU90912 Novel hum
45	562	23.2	337	6	ABO33970 Human sec

## ALIGNMENTS

RESULT 1  
AAR8058  
ID AAR8058 standard; protein; 459 AA.

AC AAR8058;  
DT 25-MAR-2003 (revised)  
DT 25-JUL-1996 (first entry)

XX Protein encoded by Mutu putative oncogene MN.

KM Mutu; endogenous; cellular component; MN; Hela cell; diagnosis;  
KM Lymphocytic Choriomeningitis virus; LCMV; putative oncogene; treatment;  
KM neoplastic; pre-neoplastic; disease; antisense therapy; antibody;  
KM vaccine; vertebrate; immunisation; carbonic anhydrase.

XX Homo sapiens.

XX	Key	Location/Qualifiers
FT	Peptide	1..37
FT	Peptide	/label= sig_peptide
FT	Peptide	36..51
FT	Peptide	/note= "anti-MN antibody epitope"
FT	Region	38..135
FT	Peptide	/note= "region homologous to collagen alpha 1 chain"
FT	Peptide	55..60
FT	Peptide	/note= "anti-MN antibody epitope"
FT	Peptide	62..67
FT	Peptide	/note= "anti-MN antibody epitope"
FT	Peptide	68..91
FT	Peptide	/note= "anti-MN antibody epitope"
FT	Peptide	127..147
FT	Peptide	/note= "anti-MN antibody epitope"
FT	Domain	136..391
FT	Peptide	/note= "carbonic anhydrase domain"
FT	Peptide	279..291
FT	Peptide	/note= "anti-MN antibody epitope"
FT	Region	414..433
FT	Region	/note= "intracellular transmembrane region"
FT	Region	434..459
FT	Peptide	/note= "intracellular C-terminus"
FT	Peptide	435..450
FT	Peptide	/note= "anti-MN antibody epitope"
XX	MO9534650-A2.	
XX	21-DEC-1995.	

PF 15-JUN-1995; 95WO-US007628.  
 XX  
 PR 15-JUN-1994; 94US-00260190.  
 PR 07-JUN-1995; 95US-00477504.  
 PR 07-JUN-1995; 95US-00481658.  
 PR 07-JUN-1995; 95US-00485049.  
 PR 07-JUN-1995; 95US-00485862.  
 PR 07-JUN-1995; 95US-00485863.  
 PR 07-JUN-1995; 95US-00486756.  
 PR 07-JUN-1995; 95US-00487077.  
 XX  
 PA (CIBA ) CIBA CORNING DIAGNOSTICS CORP.  
 PA (VIRO-) INST VIROLOGY.  
 XX  
 PI Zavađa J, Pastorekova S, Pastorek J;  
 XX WPI; 1996-049679/05.  
 DR N-PSDB; AAT09186.  
 XX  
 PT MN gene, protein and nucleic acid fragments - used as primers and probes  
 PT in the detection of MN antigens and antibodies, and in the treatment of  
 PT (pre)neoplastic disease.  
 XX  
 PS Claim 12; Fig 1; 102pp; English.  
 XX  
 CC The present sequence is encoded by the full length Mutu endogenous  
 CC cellular component, MN, cDNA clone, which was isolated from lymphocytic  
 CC choriomeningitis virus (ICMV) infected HeLa cells. Persistent ICMV, the  
 CC exogenous Mutu transmissible agent (MX), infection increases the  
 CC expression level of the MN gene. MN is a putative oncogene, and can  
 CC therefore be used in the development of probe. for the diagnosis  
 CC treatment of neoplastic (NP), or pre-NP diseases. NP diseases can be  
 CC treated using DNA antisense to MN transcribed mRNA, anti-MN protein  
 CC antibodies can be used for the diagnosis NP or pre-NP diseases and a  
 CC vaccine cong. immunogenic amounts of the MN protein can be used to  
 CC immunise a vertebrate against a NP disease associated with MN antigen  
 CC expression. (Updated on 25-MAR-2003 to correct PR field.)  
 CC  
 XX  
 SQ Sequence 459 AA;

Query Match 100.0%; Score 2424; DB 2; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-182;  
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCSPFWLPLIPAPGLITVQLLISLLIMFHPQGLPKQKESPIGGSSSGEDDPL 60  
 DB 1 MAPLCSPFWLPLIPAPGLITVQLLISLLIMFHPQGLPKQKESPIGGSSSGEDDPL 60  
 QY 61 GEDLPSEEDSPREDDPGEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 DB 61 GEDLPSEEDSPREDDPGEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 QY 121 DQSPQNNARHDKEDQSHRRYGGDPMPWRVSPACAGRQSVDRPOLAFCPLRPL 180  
 DB 121 DQSPQNNARHDKEDQSHRRYGGDPMPWRVSPACAGRQSVDRPOLAFCPLRPL 180  
 QY 181 ELIGFOLPPLBELLRNNGSVQVTLPPGLEMALGPRERFALQTLHNGAAGRPSEHT 240  
 DB 181 ELIGFOLPPLBELLRNNGSVQVTLPPGLEMALGPRERFALQTLHNGAAGRPSEHT 240  
 QY 241 VEGHRFPAEIHVHLSTAFARVDEALGRPGGLAVLAFLBEGEENSAYEQILSRLEBIA 300  
 DB 241 VEGHRFPAEIHVHLSTAFARVDEALGRPGGLAVLAFLBEGEENSAYEQILSRLEBIA 300  
 QY 301 BEGSETQVPGDISALPSDSRYFOYEGSLTPPCAGQYIVFVQVTLWSAKQHLTIS 360  
 DB 301 BEGSETQVPGDISALPSDSRYFOYEGSLTPPCAGQYIVFVQVTLWSAKQHLTIS 360  
 QY 361 DTWGPDSRLQINFRATQPLNGRVIEASPPAGVSSPRAAEVQVQNSCIAAGDITLVF 420  
 DB 361 DTWGPDSRLQINFRATQPLNGRVIEASPPAGVSSPRAAEVQVQNSCIAAGDITLVF 420  
 QY 421 GILFAVTSVAFVQMRQRHRRGTGKGVSYRPAEVAERGA 459

DB 421 GILFAVTSVAFVQMRQRHRRGTGKGVSYRPAEVAERGA 459

RESULT 2  
 AAY53228  
 ID AAY53228 standard; protein; 459 AA.  
 XX  
 AC AAY53228;  
 XX  
 DT 16-JUN-2000 (first entry)  
 XX  
 DE Human MN protein SEQ ID NO:2.  
 XX  
 KM Human; MN protein; MN gene; oncogene; carbonic anhydrase; tumour;  
 KM oncogenesis; diagnosis; neoplastic disease; cancer; carcinoma;  
 KM MN/CA IX isoenzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6027887-A.  
 XX  
 PD 22-FEB-2000.  
 XX  
 PF 24-JUN-1997; 97US-00787739.  
 XX  
 XX 21-OCT-1992; 92US-00964589.  
 PR 30-DEC-1993; 93US-00177093.  
 PR 15-JUN-1994; 94US-00260190.  
 PR 07-JUN-1995; 95US-00477504.  
 PR 07-JUN-1995; 95US-00481658.  
 PR 07-JUN-1995; 95US-00485049.  
 PR 07-JUN-1995; 95US-00485862.  
 PR 07-JUN-1995; 95US-00485863.  
 PR 07-JUN-1995; 95US-00486756.  
 PR 07-JUN-1995; 95US-00487077.  
 XX  
 PA (SLSC-) SLOVAK ACAD SCI INST VIROLOGY.  
 XX  
 PI Pastorek J, Zavađa J, Pastorekova S;  
 XX WPI; 2000-194827/17.  
 DR N-PSDB; AAA16540.  
 XX  
 PT Nucleic acid based assay for diagnosing a wide variety of  
 PT preneoplastic/neoplastic disease comprises screening for the presence of  
 PT abnormal MN gene expression in a vertebrate.  
 XX  
 PS Disclosure; Fig 1; 87pp; English.  
 XX  
 CC The present invention describes a method of screening for  
 CC preneoplastic/neoplastic disease. The method comprises: (1) determining  
 CC whether abnormal MN gene expression is present in a vertebrate; and (2)  
 CC if abnormal MN gene expression is determined to be present in the  
 CC vertebrate, determining that the vertebrate has a significant risk of  
 CC having preneoplastic/neoplastic disease. The MN gene is an oncogene and  
 CC encodes an MN protein (also referred to as MN/CA IX isoenzyme). The MN  
 CC protein is a tumour associated carbonic anhydrase isoenzyme. The method  
 CC is used for detecting a wide variety of preneoplastic/neoplastic diseases  
 CC in a vertebrate, preferably a human. The disease detected is mammary,  
 CC bladder, renal, urinary tract, ovarian, uterine, cervical, endometrial,  
 CC vaginal, vulval, prostate, liver, lung, skin, thyroid, pancreatic,  
 CC testicular, brain, head and neck, mesodermal, gallbladder, rectal,  
 CC duodenal, jejunal, ileal, gastric, pancreatic duct, liver duct, gastric  
 CC mucosa, gallbladder epithelium, small intestinal mucosa, colorectal  
 CC mucosa, pancreatic duct epithelium or liver duct epithelium  
 CC preneoplastic/neoplastic disease. AAA16540 to AAA16617 and AAY53228 to  
 CC AAY53245 represent sequences used in the exemplification of the present  
 CC invention  
 XX  
 SQ Sequence 459 AA;

Query Match 100.0%; Score 2424; DB 3; Length 459;



Best Local Similarity 100.0%; Pred. No. 1,1e-182; Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAPLCSPWMLPLIPAPAGLTIVQLLSLLLMFPHQRLPMQDSPLGGSSGSDPL 60
DB 1 MAPLCSPWMLPLIPAPAGLTIVQLLSLLLMFPHQRLPMQDSPLGGSSGSDPL 60
QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPVEXKSEEGSLKLEDLPTVEAPG 120
DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPVEXKSEEGSLKLEDLPTVEAPG 120
QY 121 DPQEPQNNARHDKGDDQSHMRVGGDPMPRVSPACAGFQSPVDIRPOLAFCALRPL 180
DB 121 DPQEPQNNARHDKGDDQSHMRVGGDPMPRVSPACAGFQSPVDIRPOLAFCALRPL 180
QY 181 ELLGFQPLPPELRLNNGHSVOLTPPGLEMAFGREYRALQHLHMGAAAGRGSEHT 240
DB 181 ELLGFQPLPPELRLNNGHSVOLTPPGLEMAFGREYRALQHLHMGAAAGRGSEHT 240
QY 241 VEGHRPFAEIHVHVLSTAFARVDEALGRPGGLAVLAFLBEGPEENSAYEQLSLREIA 300
DB 241 VEGHRPFAEIHVHVLSTAFARVDEALGRPGGLAVLAFLBEGPEENSAYEQLSLREIA 300
QY 301 EEGSETQVPGDLISALLPDSFSRYFOYEGSLTTPCAQGVITVFNQVMSAKQLHTLS 360
DB 301 EEGSETQVPGDLISALLPDSFSRYFOYEGSLTTPCAQGVITVFNQVMSAKQLHTLS 360
QY 361 DTLMGPDSSLQNFRAQPLNGRVLEASFPAGVDSPPRAAPVOLNSCLAAGDILALVF 420
DB 361 DTLMGPDSSLQNFRAQPLNGRVLEASFPAGVDSPPRAAPVOLNSCLAAGDILALVF 420
QY 421 GLFPAVTSVAFLVQMRQRHGRGTGKGVSPRAEVAETGA 459
DB 421 GLFPAVTSVAFLVQMRQRHGRGTGKGVSPRAEVAETGA 459

```

## RESULT 3

AAB03005 ID AAB03005 standard; protein; 459 AA.

XX AAB03005;

DT 25-SEP-2000 (first entry)

DE Human MN protein.

XX MN protein; tumour associated cell adhesion molecule; oncoprotein;

KM proteoglycan domain; PG domain; carbonic anhydrase; Ca domain;

XX abnormal expression; neoplastic disease; cancer; gene therapy.

OS Homo sapiens.

XX WO200024913-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US024879.

XX 23-OCT-1998; 98US-0017776.

XX 23-OCT-1998; 98US-0017815.

XX (FAB) BAYER CORP.

XX (VIRCO) INST VIROLOGY.

XX Zavada J, Pastorekova S, Pastorek J;

XX WPI; 2000-350752/30.

XX N-PSDB; AAA52459, AAA52462.

XX A molecule which specifically binds to a site on MN protein (oncoprotein) and prevents adhesion of vertebrate cells to the protein, useful for treating preneoplastic or neoplastic diseases such as cancer.

PS Example 1; Fig 1A-C; 154bp; English.

CC The invention relates to the inhibition of cell adhesion mediated by the  
 CC MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G250  
 CC protein). The MN protein is a tumour-associated adhesion molecule which  
 CC comprises a proteoglycan-like (PG) domain (AAB03017) which contains the  
 CC protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018).  
 CC Abnormal expression of the MN protein is associated with tumorigenicity.  
 CC The invention encompasses molecules (e.g., proteins and peptides) which  
 CC which specifically bind to a site on the MN protein, thereby preventing  
 CC adhesion of vertebrate cells to the protein in a cell adhesion assay. It  
 CC also encompasses MN proteins or MN protein fragments which can be added  
 CC to the extracellular environment to prevent the adhesion of vertebrate  
 CC cells to each other. The invention also relates to the identification of  
 CC the binding site of the MN protein and to a method of identifying a site  
 CC on an MN protein to which cells adhere, comprising testing a series of  
 CC overlapping peptides from the protein in a cell adhesion assay. The  
 CC invention encompasses a vector comprising an expression control sequence  
 CC operatively linked to a nucleic acid encoding the variable domains of a  
 CC MN-specific antibody, where the domains are separated by a flexible  
 CC linker peptide (AAB03035) and the vector inhibits the growth of a  
 CC vertebrate preneoplastic or neoplastic cell that abnormally expresses MN  
 CC protein. The invention also encompasses a vector comprising a nucleic  
 CC acid encoding a cytotoxic protein or peptide operatively linked to the MN  
 CC gene promoter, which inhibits the growth of a vertebrate preneoplastic or  
 CC neoplastic cell. Also claimed is a repressor complex that binds to the MN  
 CC gene promoter (AAA52473). MN proteins and peptides, MN-binding proteins  
 CC and peptides, and expression vectors encoding such proteins and peptides  
 CC are useful for treating patients with preneoplastic or neoplastic disease  
 CC (e.g., cancers) associated with or characterised by abnormal MN  
 CC expression. The present sequence represents the human MN protein.

XX Sequence 459 AA:

Query Match 100.0%; Score 2424; DB 3; Length 459;

Best Local Similarity 100.0%; Pred. No. 1,1e-182; Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAPLCSPWMLPLIPAPAGLTIVQLLSLLLMFPHQRLPMQDSPLGGSSGSDPL 60
DB 1 MAPLCSPWMLPLIPAPAGLTIVQLLSLLLMFPHQRLPMQDSPLGGSSGSDPL 60
QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPVEXKSEEGSLKLEDLPTVEAPG 120
DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPVEXKSEEGSLKLEDLPTVEAPG 120
QY 121 DPQEPQNNARHDKGDDQSHMRVGGDPMPRVSPACAGFQSPVDIRPOLAFCALRPL 180
DB 121 DPQEPQNNARHDKGDDQSHMRVGGDPMPRVSPACAGFQSPVDIRPOLAFCALRPL 180
QY 181 ELLGFQPLPPELRLNNGHSVOLTPPGLEMAFGREYRALQHLHMGAAAGRGSEHT 240
DB 181 ELLGFQPLPPELRLNNGHSVOLTPPGLEMAFGREYRALQHLHMGAAAGRGSEHT 240
QY 241 VEGHRPFAEIHVHVLSTAFARVDEALGRPGGLAVLAFLBEGPEENSAYEQLSLREIA 300
DB 241 VEGHRPFAEIHVHVLSTAFARVDEALGRPGGLAVLAFLBEGPEENSAYEQLSLREIA 300
QY 301 EEGSETQVPGDLISALLPDSFSRYFOYEGSLTTPCAQGVITVFNQVMSAKQLHTLS 360
DB 301 EEGSETQVPGDLISALLPDSFSRYFOYEGSLTTPCAQGVITVFNQVMSAKQLHTLS 360
QY 361 DTLMGPDSSLQNFRAQPLNGRVLEASFPAGVDSPPRAAPVOLNSCLAAGDILALVF 420
DB 361 DTLMGPDSSLQNFRAQPLNGRVLEASFPAGVDSPPRAAPVOLNSCLAAGDILALVF 420
QY 421 GLFPAVTSVAFLVQMRQRHGRGTGKGVSPRAEVAETGA 459
DB 421 GLFPAVTSVAFLVQMRQRHGRGTGKGVSPRAEVAETGA 459

```

## RESULT 4

AAE17175

ID AAE17175 standard; protein; 459 AA.  
 AC AAE17175;  
 DT 18-APR-2002 (first entry)  
 DE Human RCC-associated antigen, G250 protein.  
 XX  
 DE Human; immune response; immunogenic tumour; renal cell carcinoma;  
 KW RCC-associated antigen; G250 protein; immunotherapy; gene therapy;  
 KW cancer; cytostatic; immunostimulant.  
 OS Homo sapiens.  
 PN WO200198363-A2.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 20-JUN-2001; 2001WO-NL000461.  
 XX  
 PR 20-JUN-2000; 2000US-0212669P.  
 XX  
 PA (UYN1-) UNIV Nijmegen.  
 XX  
 PI Vissers JLM, De Vries JCM, Oosterwijk E, Figdor CG, Adema GJ;  
 DR WPI; 2002-139784/18.  
 XX  
 PT New peptides useful in preparing a composition for the treatment of  
 PT cancer, including renal cell carcinoma, cancer of the kidney, prostate,  
 PT head, neck, or gastrointestinal tract, and in eliciting immune response.  
 XX  
 PS Claim 1; Page 35-36; 39pp; English.  
 XX  
 CC The patent discloses peptides which can be used to elicit an immune  
 CC response against a tumour, specifically against an immunogenic tumour.  
 CC The peptides are derived from the amino acid sequence of renal cell  
 CC carcinoma (RCC)-associated antigen, G250 protein. They are used in  
 CC immunotherapy of tumours, in particular renal cell carcinomas. Sequences  
 CC of the invention are useful in preparing a composition for the treatment  
 CC of cancer, including renal cell carcinoma, cancer of the kidney, head,  
 CC prostate, neck, stomach, colon, gastrointestinal tract and bladder. They  
 CC are also used for treating immunogenic tumours. Nucleic acid sequences  
 CC encoding the peptides of the invention are used in gene therapy. The  
 CC present sequence is human RCC-associated antigen, G250 protein  
 XX  
 SO Sequence 459 AA.

Query Match 100.0%; Score 2424; DB 5; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-182;  
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCSPFWLPLIPAPGLTVQLLSILLMPVAPQLPFMOEDSPAGGSSGEDDPL 60  
 DB 1 MAPLCSPFWLPLIPAPGLTVQLLSILLMPVAPQLPFMOEDSPAGGSSGEDDPL 60  
 QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 QY 121 DPOEPONNARDEKGDQSHMRVGGDPFMRVSPACAGFQSPVDIRPDLAFCALAPL 180  
 DB 121 DPOEPONNARDEKGDQSHMRVGGDPFMRVSPACAGFQSPVDIRPDLAFCALAPL 180  
 QY 181 ELIGFOLPLPELRLNNGHVSQVLTLPFGLEMALGGRYRALQLHLHGAAGRGSSHT 240  
 DB 181 ELIGFOLPLPELRLNNGHVSQVLTLPFGLEMALGGRYRALQLHLHGAAGRGSSHT 240  
 QY 241 VEGHRFPAEIHWLSTAFARVDEALGRPGGLAVLAFAEEGPEENSAVEQLSLLEBIA 300  
 DB 241 VEGHRFPAEIHWLSTAFARVDEALGRPGGLAVLAFAEEGPEENSAVEQLSLLEBIA 300  
 QY 301 EEGSETQVPGJLIDISALLPSDPSRYFQYEGSLTTPCAQGVIMTVFNQYMLSAKQLHTLS 360

DB 301 EEGSETQVPGJLIDISALLPSDPSRYFQYEGSLTTPCAQGVIMTVFNQYMLSAKQLHTLS 360  
 QY 361 DTLMQPGDSRLQNLNFRATQPLNGRVIEASFAGVDSPPRAAEVQINSCLIAADILALVF 420  
 DB 361 DTLMQPGDSRLQNLNFRATQPLNGRVIEASFAGVDSPPRAAEVQINSCLIAADILALVF 420  
 QY 421 GILFAVTSVAFVLVQMRORRGTGKGVSYPPAEVAETGA 459  
 DB 421 GILFAVTSVAFVLVQMRORRGTGKGVSYPPAEVAETGA 459

RESULT 5  
 ABR58596  
 ID ABR58596 standard; protein; 459 AA.  
 AC ABR58596;  
 DT 09-JUL-2003 (first entry)  
 DE Human cancer related protein SEQ ID NO:253.  
 XX  
 DE Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;  
 KW heart disease; atherosclerosis; endometriosis.  
 OS Homo sapiens.  
 PN WO2003025138-A2.  
 XX  
 PD 27-MAR-2003.  
 XX  
 PF 17-SEP-2002; 2002WO-US029560.  
 XX  
 PR 17-SEP-2001; 2001US-0323469P.  
 PR 20-SEP-2001; 2001US-0323887P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 08-FEB-2002; 2002US-0358145P.  
 PR 08-FEB-2002; 2002US-0355257P.  
 PR 12-APR-2002; 2002US-0372246P.  
 XX  
 PA (BOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 PI Afar D, Aziz N, Glsh KC, Hevezi PA, Mack DH, Wilson KE;  
 PI Ziolkni A;  
 DT WPI; 2003-354600/33.  
 DR N-PSDB; ACC72730.  
 XX  
 PT New genes that are up-regulated or down-regulated in cancers, useful as  
 PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as  
 PT therapeutic targets for screening drugs for treating these diseases.  
 XX  
 PS Claim 12; Page 744; 767pp; English.  
 XX  
 CC The present invention describes an isolated nucleic acid molecule, which  
 CC comprises the sequence of any of the genes that are up-regulated or down-  
 CC regulated in specific cancers (e.g. about 1031 genes up-regulated in  
 CC acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer  
 CC related gene nucleotide sequences which encode the proteins given in  
 CC ABR5521 to ABR58709. Also described: (1) determining the presence or  
 CC absence of a pathological cell in a patient; (2) an expression vector  
 CC comprising a nucleic acid molecule described above; (3) a host cell  
 CC comprising the vector; (4) an isolated polypeptide, which is encoded by  
 CC the nucleic acid; (5) an antibody that specifically binds the polypeptide  
 CC of (4); (6) specifically targeting a compound to a pathological cell in a  
 CC patient by administering to the patient the antibody above; and (7) a  
 CC drug screening assay. The nucleic acid is useful as diagnostic markers or  
 CC therapeutic targets. In particular, the nucleic acid is useful for  
 CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,  
 CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,  
 CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,  
 CC atherosclerosis and endometriosis. The nucleic acid is also useful in  
 CC drug screening, particularly for identifying agents for treating these

CC pathologies  
XX Sequence 459 AA;  
SQ

Query Match 100.0%; Score 2424; DB 6; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1,1e-182;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCSPWMLPPLIPAPAGLTVOQLLSLLILMPVHPQRLPRMOBDSPIGGSSGSEDDPL 60  
DB 1 MAPLCSPWMLPPLIPAPAGLTVOQLLSLLILMPVHPQRLPRMOBDSPIGGSSGSEDDPL 60  
QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
QY 121 DPOEPQNNARHDKGDDQSHWRVYGGDPPMPRVSPACAGRFQSPVDIRPOLAFCPALRPL 180  
DB 121 DPOEPQNNARHDKGDDQSHWRVYGGDPPMPRVSPACAGRFQSPVDIRPOLAFCPALRPL 180  
QY 181 ELIGFOLPLPELRLNNGHSVQTLPPGLEMALGPGRYRALQLHMGAGRPGSEHT 240  
DB 181 ELIGFOLPLPELRLNNGHSVQTLPPGLEMALGPGRYRALQLHMGAGRPGSEHT 240  
QY 241 VEGHFPFAIHHVHLSTAFARVDALGRPGGLAVLAFLKEGPEENSAYEQLLSRLEEA 300  
DB 241 VEGHFPFAIHHVHLSTAFARVDALGRPGGLAVLAFLKEGPEENSAYEQLLSRLEEA 300  
QY 301 EEGSETQVPGDLISALLPSDFSRFYQEGSLTTPCAQGVITVFNQVWLSAKOHLTIS 360  
DB 301 EEGSETQVPGDLISALLPSDFSRFYQEGSLTTPCAQGVITVFNQVWLSAKOHLTIS 360  
QY 361 DTLMGPGDSRLQNFRAQPLNGRVIEASPPAGVDSPPRAAEVQVQNSCLAAGDILALVF 420  
DB 361 DTLMGPGDSRLQNFRAQPLNGRVIEASPPAGVDSPPRAAEVQVQNSCLAAGDILALVF 420  
QY 421 GLFPAVTSVAFVQMRQHRGRTKGVSYPRAEVAETGA 459  
DB 421 GLFPAVTSVAFVQMRQHRGRTKGVSYPRAEVAETGA 459

RESULT 6  
ABP97744  
ID ABP97744 standard; protein; 459 AA.  
XX  
AC ABP97744;  
XX  
DT 28-MAY-2003 (first entry)  
XX  
DE Amino acid sequence of human CA9 polypeptide.  
XX  
KM Human; angiogenesis; wound healing; retinopathy; ischemia; inflammation;  
KM microvasculopathy; bone healing; skin inflammation; HOG3; HOG8; HOG16;  
KM follicular development; CA9; HXB; IGFBP5; HFAFP; STC1; mlg-6; SSR4;  
KM cancer.  
XX  
OS Homo sapiens.  
XX  
PN MO2003010205-AL.  
XX  
PD 06-FEB-2003.  
XX  
PF 26-JUL-2002; 2002MO-US023796.  
XX  
PR 26-JUL-2001; 2001US-0307600P.  
XX  
PR 24-JUL-2002; 2002US-00201642.  
XX  
PA (UYDU-) UNIV DUKE MEDICAL CENT.  
XX  
PI Riggins GJ, Lal A;  
XX  
DR WPI; 2003-239423/23.  
DR N-PSDB; AB277284.

XX Inhibiting angiogenesis for treating wound healing, retinopathy,  
PT ischemia, inflammation, microvasculopathy, bone healing, skin  
PT inflammation or follicular development by providing to a subject an  
PT antisense polynucleotide.  
XX  
PS Claim 4; Page 47-48; 66p; English.  
XX  
CC The present sequence is a human CA9 polypeptide. It is used in the method  
CC of the invention. The specification describes a method modulating  
CC angiogenesis associated with wound healing, retinopathy, ischemia,  
CC inflammation, microvasculopathy, bone healing, skin inflammation or  
CC follicular development. The method comprises providing to a subject HOG3,  
CC HOG8, HOG18, CA9, HXB, IGFBP5, HFAFP, STC1, mlg-6 or SSR4. The methods,  
CC antisense polynucleotides, polypeptides and antibodies are useful for  
CC treating wound healing, retinopathy, ischemia, inflammation,  
CC microvasculopathy, bone healing, skin inflammation or follicular  
CC development, or cancer such as breast, colon or lung cancer, or  
CC glioblastoma  
XX  
SQ Sequence 459 AA;

Query Match 100.0%; Score 2424; DB 6; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1,1e-182;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCSPWMLPPLIPAPAGLTVOQLLSLLILMPVHPQRLPRMOBDSPIGGSSGSEDDPL 60  
DB 1 MAPLCSPWMLPPLIPAPAGLTVOQLLSLLILMPVHPQRLPRMOBDSPIGGSSGSEDDPL 60  
QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
QY 121 DPOEPQNNARHDKGDDQSHWRVYGGDPPMPRVSPACAGRFQSPVDIRPOLAFCPALRPL 180  
DB 121 DPOEPQNNARHDKGDDQSHWRVYGGDPPMPRVSPACAGRFQSPVDIRPOLAFCPALRPL 180  
QY 181 ELIGFOLPLPELRLNNGHSVQTLPPGLEMALGPGRYRALQLHMGAGRPGSEHT 240  
DB 181 ELIGFOLPLPELRLNNGHSVQTLPPGLEMALGPGRYRALQLHMGAGRPGSEHT 240  
QY 241 VEGHFPFAIHHVHLSTAFARVDALGRPGGLAVLAFLKEGPEENSAYEQLLSRLEEA 300  
DB 241 VEGHFPFAIHHVHLSTAFARVDALGRPGGLAVLAFLKEGPEENSAYEQLLSRLEEA 300  
QY 301 EEGSETQVPGDLISALLPSDFSRFYQEGSLTTPCAQGVITVFNQVWLSAKOHLTIS 360  
DB 301 EEGSETQVPGDLISALLPSDFSRFYQEGSLTTPCAQGVITVFNQVWLSAKOHLTIS 360  
QY 361 DTLMGPGDSRLQNFRAQPLNGRVIEASPPAGVDSPPRAAEVQVQNSCLAAGDILALVF 420  
DB 361 DTLMGPGDSRLQNFRAQPLNGRVIEASPPAGVDSPPRAAEVQVQNSCLAAGDILALVF 420  
QY 421 GLFPAVTSVAFVQMRQHRGRTKGVSYPRAEVAETGA 459  
DB 421 GLFPAVTSVAFVQMRQHRGRTKGVSYPRAEVAETGA 459

RESULT 7  
ABUS6656  
ID ABUS6656 standard; protein; 459 AA.  
XX  
AC ABUS6656;  
XX  
DT 02-APR-2003 (first entry)  
XX  
DE Lung cancer-associated polypeptide #249.  
XX  
KM Lung cancer-associated polypeptide; cytostatic; emphysema;  
KM antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
KM small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
KM chronic obstructive pulmonary disease; hypersensitivity pneumonitis;

interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
 XX Unidentified.  
 CS  
 XX WO200286443-A2.  
 EN  
 XX 31-OCT-2002.  
 PD  
 XX 18-APR-2002; 2002WO-US012476.  
 PF  
 XX 18-APR-2001; 2001US-0284770P.  
 PR 10-MAY-2001; 2001US-0280492P.  
 PR 09-NOV-2001; 2001US-039245P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 29-NOV-2001; 2001US-034370P.  
 PR 12-APR-2002; 2002US-0372246P.  
 XX  
 PA (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 PI Aziz N, Murray R;  
 XX WPI; 2003-093161/08.  
 DR N-PSDB; ABX76385.  
 DR  
 XX  
 PT Detecting a lung cancer-associated transcript in a cell from a patient  
 PT for treating lung cancer, by contacting a biological sample from the  
 PT patient with a polynucleotide that exhibits increased or decreased  
 PT expression in lung cancer.  
 PT  
 XX  
 PS Claim 27; Page 380; 453pp; English.  
 XX  
 XX The invention relates to a method for detecting a lung cancer-associated  
 CC transcript in a cell from a patient, comprising contacting a biological  
 CC sample from the patient with a polynucleotide that selectively hybridizes  
 CC to a sequence that is at least 80 % identical to a gene that exhibits  
 CC increased or decreased expression in lung cancer samples. Lung cancer-  
 CC associated polynucleotides and polypeptides are used for identifying a  
 CC compound that modulates a lung cancer-associated polypeptide, for  
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung  
 CC cancer in a patient and for treating a mammal having lung cancer by  
 CC administering a modulatory compound identified. The methods are useful  
 CC for treating lung cancer, such as small cell lung cancer, non-small cell  
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful  
 CC for diagnostic purposes and as targets for screening for therapeutic  
 CC compounds that modulate lung cancer, such as antibodies. Sequences of  
 CC ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the  
 CC invention  
 CC  
 XX  
 SQ Sequence 459 AA;  
 Query Match 100.0%; Score 2424; DB 6; Length 459;  
 Best local similarity 100.0%; Pred. No. 1, 1e-182;  
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAPLCSPWLPPLIPAPAPGLTVQLLSTLLMPVAPQALPQMDSPLGGSSGEDPPL 60  
 DB 1 MAPLCSPWLPPLIPAPAPGLTVQLLSTLLMPVAPQALPQMDSPLGGSSGEDPPL 60  
 QY 61 GEEPLPSEEDSPREDDPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 DB 61 GEEPLPSEEDSPREDDPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 QY 121 DPOBPONNARDEKGDQSHMRXGDPWPWPVRS PACAGSFQSPVDIRPOLAFCALRPL 180  
 DB 121 DPOBPONNARDEKGDQSHMRXGDPWPWPVRS PACAGSFQSPVDIRPOLAFCALRPL 180  
 QY 181 ELIGFOPLPLPELRLRNHGSVQTLPLGLSMALGGRERYRALQHLHMGAGRGSGSHT 240  
 DB 181 ELIGFOPLPLPELRLRNHGSVQTLPLGLSMALGGRERYRALQHLHMGAGRGSGSHT 240

QY 241 VEGHRRPAEIHVHLSTAFARVDEALGRPGLAVALAELEGPEENSAYEQLSRLEEA 300  
 DB 241 VEGHRRPAEIHVHLSTAFARVDEALGRPGLAVALAELEGPEENSAYEQLSRLEEA 300  
 QY 301 EBSGETQVPEGLDISALPSDPSRRYQYFGSLTTPCAQGVITWVFNQTMLSAQVHTLS 360  
 DB 301 EBSGETQVPEGLDISALPSDPSRRYQYFGSLTTPCAQGVITWVFNQTMLSAQVHTLS 360  
 QY 361 DTLMGPGDSRLQINRATOPUNGRVTEASFPAGVDSFPRAEPVQLNSCLAAGITLAVF 420  
 DB 361 DTLMGPGDSRLQINRATOPUNGRVTEASFPAGVDSFPRAEPVQLNSCLAAGITLAVF 420  
 QY 421 GLIFAVTSVAFIVQMRQRRGTGCVSYRAEVAETGA 459  
 DB 421 GLIFAVTSVAFIVQMRQRRGTGCVSYRAEVAETGA 459  
 RESULT 8  
 ID ABUS6408 standard; protein; 459 AA.  
 XX  
 AC ABUS6408;  
 XX  
 DT 02-APR-2003 (first entry)  
 DE Lung cancer-associated polypeptide #1.  
 XX  
 XX Lung cancer-associated polypeptide; cytostatic; emphysema;  
 KW antiinflammatory; antiaesthetic; non-small cell lung cancer; atelectasis;  
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
 OS Unidentified.  
 XX  
 PN WO200286443-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 18-APR-2002; 2002WO-US012476.  
 PR 18-APR-2001; 2001US-0284770P.  
 PR 10-MAY-2001; 2001US-0290492P.  
 PR 09-NOV-2001; 2001US-0339245P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 29-NOV-2001; 2001US-034370P.  
 PR 12-APR-2002; 2002US-0372246P.  
 XX  
 PA (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 PI Aziz N, Murray R;  
 XX WPI; 2003-093161/08.  
 DR N-PSDB; ABX76124.  
 DR  
 XX  
 PT Detecting a lung cancer-associated transcript in a cell from a patient  
 PT for treating lung cancer, by contacting a biological sample from the  
 PT patient with a polynucleotide that exhibits increased or decreased  
 PT expression in lung cancer.  
 PT  
 XX  
 PS Claim 27; Page 189; 453pp; English.  
 XX  
 XX The invention relates to a method for detecting a lung cancer-associated  
 CC transcript in a cell from a patient, comprising contacting a biological  
 CC sample from the patient with a polynucleotide that selectively hybridizes  
 CC to a sequence that is at least 80 % identical to a gene that exhibits  
 CC increased or decreased expression in lung cancer samples. Lung cancer-  
 CC associated polynucleotides and polypeptides are used for identifying a  
 CC compound that modulates a lung cancer-associated polypeptide, for  
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung  
 CC cancer in a patient and for treating a mammal having lung cancer by  
 CC administering a modulatory compound identified. The methods are useful  
 CC for treating lung cancer, such as small cell lung cancer, non-small cell



Db 507 TLMPGDSRLQNLFRATQPLNGRVIKSPAGVDSPPRAAEVQNSCLAADILAVFG 566

QY 422 LLFAVTSVAFVQMRQHRGCTKGVSYPRAEVAEFGA 459

Db 567 LLFAVTSVAFVQMRQHRGCTKGVSYPRAEVAEFGA 604

RESULT 10  
AAV53245  
ID AAV53245 standard; protein; 377 AA.

AC AAV53245;

XX 16-JUN-2000 (first entry)

XX MN protein extracellular domain SEQ ID NO:87.

DE Human; MN protein; MN gene; oncogene; carbonic anhydrase; tumour;  
KW oncogenesis; diagnosis; neoplastic disease; cancer; carcinoma;  
KM MN/CA IX isoenzyme.

XX Homo sapiens.

OS US6027887-A.

PN 22-FEB-2000.

PD 24-JAN-1997; 97US-00787739.

XX 21-OCT-1992; 92US-00964589.  
PR 30-DEC-1993; 93US-00177093.  
PR 15-JUN-1994; 94US-00260190.  
PR 07-JUN-1995; 95US-00477504.  
PR 07-JUN-1995; 95US-00481658.  
PR 07-JUN-1995; 95US-00485049.  
PR 07-JUN-1995; 95US-00485862.  
PR 07-JUN-1995; 95US-00485863.  
PR 07-JUN-1995; 95US-00486756.  
PR 07-JUN-1995; 95US-00487077.

XX (SLSC-) SLOVAK ACAD SCI INST VIROLOGY.

XX Pastorek J, Zavada J, Pastorekova S;  
PI WPI; 2000-194827/17.

XX Nucleic acid based assay for diagnosing a wide variety of  
PT preneoplastic/neoplastic disease comprises screening for the presence of  
PT abnormal MN gene expression in a vertebrate.

XX Disclosure; Col 16; 87pp; English.

XX The present invention describes a method of screening for  
XX preneoplastic/neoplastic disease. The method comprises: (1) determining  
XX whether abnormal MN gene expression is present in a vertebrate; and (2)  
XX if abnormal MN gene expression is determined to be present in the  
XX vertebrate, determining that the vertebrate has a significant risk of  
XX having preneoplastic/neoplastic disease. The MN gene is an oncogene and  
XX encodes an MN protein (also referred to as MN/CA IX isoenzyme). The MN  
XX protein is a tumour associated carbonic anhydrase. The method  
XX is used for detecting a wide variety of preneoplastic/neoplastic diseases  
XX in a vertebrate, preferably a human. The disease detected is mammary,  
XX bladder, renal, urinary tract, ovarian, uterine, cervical, endometrial,  
XX vaginal, vulval, prostate, liver, lung, skin, thyroid, pancreatic,  
XX testicular, brain, head and neck, mesodermal, gallbladder, rectal,  
XX duodenal, jejunal, ileal, gastric, pancreatic duct, liver duct, gastric  
XX mucosa, gallbladder epithelium, small intestinal mucosa, colorectal  
XX mucosa, pancreatic duct epithelium or liver duct epithelium  
XX preneoplastic/neoplastic disease. AAV5340 to AAV561 and AAV53228 to  
XX AAV53245 represent sequences used in the exemplification of the present  
XX invention  
XX Sequence 377 AA;

Query Match 82.8%; Score 2007; DB 3; Length 377;  
Best Local Similarity 100.0%; Pred. No. 6; 9e-150; Mismatches 0; Gaps 0;  
Matches 377; Conservative 0; Indels 0;

QY 38 ORLPRMQEDSPVGGSGEDDPLGEDDPLSEEDSPREEDPGEEDLPGEEDLPGEEDLP 97

Db 1 ORLPRMQEDSPVGGSGEDDPLGEDDPLSEEDSPREEDPGEEDLPGEEDLPGEEDLP 60

QY 98 VPKSPSEBSLLEBULPYEAPGDPQEPQNNANRDKKEDDSSHRYGGDPPMPVSPACA 157

Db 61 VPKSPSEBSLLEBULPYEAPGDPQEPQNNANRDKKEDDSSHRYGGDPPMPVSPACA 120

QY 158 GRFQSPVDIRPOLAFCALRPLELGFQLPPLPELRNRNNGHVQQLTLPGLSMALGPG 217

Db 121 GRFQSPVDIRPOLAFCALRPLELGFQLPPLPELRNRNNGHVQQLTLPGLSMALGPG 180

QY 218 REYRALQLHLMWGAAGRQSGETHVGHFPPELIVVHISTAFVDEALGRPGSLAVLAA 277

Db 181 REYRALQLHLMWGAAGRQSGETHVGHFPPELIVVHISTAFVDEALGRPGSLAVLAA 240

QY 278 FLEBGPENSAYEQLLSRLERIAEGSETOVPGDISALPSPDSRYFOYEGSLTTPPCA 337

Db 241 FLEBGPENSAYEQLLSRLERIAEGSETOVPGDISALPSPDSRYFOYEGSLTTPPCA 300

QY 338 QGVIVTFNQVTVMSAKQLHTLSDTLWPGDSRLQNLFRATQPLNGRVIKSPAGVDS 397

Db 301 QGVIVTFNQVTVMSAKQLHTLSDTLWPGDSRLQNLFRATQPLNGRVIKSPAGVDS 360

QY 398 PRAAEVQNSCLAAGD 414

Db 361 PRAAEVQNSCLAAGD 377

RESULT 11  
AAB03021  
ID AAB03021 standard; protein; 377 AA.

XX AAB03021;

AC 25-SEP-2000 (first entry)

XX Mature human MN protein N-terminal extracellular domain.

XX MN protein; tumour associated cell adhesion molecule; oncoprotein;  
KW proteoglycan domain; PG domain; carbonic anhydrase; CA domain;  
KM abnormal expression; neoplastic disease; cancer; gene therapy.

XX Homo sapiens.

XX WO200024913-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US024879.

XX 23-OCT-1998; 98US-00177776.  
PR 23-OCT-1998; 98US-00178115.

XX (FAP) BAYER CORP.  
PA (VIRO-) INST VIROLOGY.

XX Zavada J, Pastorekova S, Pastorek J;  
PI WPI; 2000-150752/30.

XX A molecule which specifically binds to a site on MN protein (oncoprotein)  
PT and prevents adhesion of vertebrate cells to the protein, useful for  
PT creating preneoplastic or neoplastic diseases such as cancer.

XX Disclosure; Page 137-138; 154pp; English.

XX The invention relates to the inhibition of cell adhesion mediated by the

CC MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G250  
 CC protein). The MN protein is a tumour-associated adhesion molecule which  
 CC comprises a proteoglycan-like (PG) domain (AAB03017) which contains the  
 CC protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018).  
 CC Abnormal expression of the MN protein is associated with tumorigenicity.  
 CC The invention encompasses molecules (e.g., proteins and peptides) which  
 CC which specifically bind to a site on the MN protein, thereby preventing  
 CC adhesion of vertebrate cells to the protein in a cell adhesion assay. It  
 CC also encompasses MN proteins or MN protein fragments which can be added  
 CC to the extracellular environment to prevent the adhesion of vertebrate  
 CC cells to each other. The invention also relates to the identification of  
 CC the binding site of the MN protein and to a method of identifying a site  
 CC on an MN protein to which cells adhere, comprising testing a series of  
 CC overlapping peptides from the protein in a cell adhesion assay. The  
 CC invention encompasses a vector comprising an expression control sequence  
 CC operatively linked to a nucleic acid encoding the variable domains of a  
 CC MN-specific antibody, where the domains are separated by a flexible  
 CC linker peptide (AAB03035) and the vector inhibits the growth of a  
 CC vertebrate preneoplastic or neoplastic cell that abnormally expresses MN  
 CC protein. The invention also encompasses a vector comprising a nucleic  
 CC acid encoding a cytotoxic protein or peptide operatively linked to the MN  
 CC gene promoter, which inhibits the growth of a vertebrate preneoplastic or  
 CC neoplastic cell. Also claimed is a repressor complex that binds to the MN  
 CC gene promoter (AAB02473). MN proteins and peptides, MN-binding proteins  
 CC and peptides, and expression vectors encoding such proteins and peptides  
 CC are useful for treating patients with preneoplastic or neoplastic disease  
 CC (e.g., cancers) associated with or characterized by abnormal MN  
 CC expression. Sequences AAB03006, AAB03015-B03022, AAB03023-B03031,  
 CC AAB03036 and AAB03059-B03060 represent specific domains or regions of the  
 CC human MN protein (AAB03005) referred to in the invention  
 CC  
 XX Sequence 377 AA;  
 SQ

Query Match 82.8%; Score 2007; DB 3; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-150;  
 Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 QRLPRMOEDSPGCGSSGDDPLGEEDLPSEEDSPREDDPGEEDLPGEEDLP 97  
 DB 1 QRLPRMOEDSPGCGSSGDDPLGEEDLPSEEDSPREDDPGEEDLPGEEDLP 60  
 QY 98 VKKSEBEGSLKLEDDPTVEAPDPOEPONNAHDKXGDSQHWYGGDPMPRYSRPA 157  
 DB 61 VKKSEBEGSLKLEDDPTVEAPDPOEPONNAHDKXGDSQHWYGGDPMPRYSRPA 120  
 QY 158 GRFQSPVDIRPOLAFCPALRPLELIGFQLPPLPELRLRNNGHSVQTLTPGLEMAJSPG 217  
 DB 121 GRFQSPVDIRPOLAFCPALRPLELIGFQLPPLPELRLRNNGHSVQTLTPGLEMAJSPG 180  
 QY 218 REVRALQHLHMGAAARPGSEHTEVEGHRPAETIYVHLSTARVDEALGRPGGLAVYAA 277  
 DB 181 REVRALQHLHMGAAARPGSEHTEVEGHRPAETIYVHLSTARVDEALGRPGGLAVYAA 240  
 QY 278 FLEEGPENSAVYQLLSRLLEIAEGSETOVPGDLISALPSDFSRFYQYSGSLTPPCA 337  
 DB 241 FLEEGPENSAVYQLLSRLLEIAEGSETOVPGDLISALPSDFSRFYQYSGSLTPPCA 300  
 QY 338 QGVIMTVENOTVWLSAKOJHTLSDTLWGPDSRLQINPANOPLNGRIEYSPFAGVDS 397  
 DB 301 QGVIMTVENOTVWLSAKOJHTLSDTLWGPDSRLQINPANOPLNGRIEYSPFAGVDS 360  
 QY 398 PRAAEPYQLNSCLAAGD 414  
 DB 361 PRAAEPYQLNSCLAAGD 377

RESULT 12  
 AAR41746 standard; protein; 429 AA.  
 ID AAR41746;  
 AC AAR41746;  
 XX  
 XX 25-MAR-2003 (revised)

DT 25-MAR-1994 (first entry)  
 XX  
 XX MN protein.  
 DE MN protein.  
 XX  
 XX MN; endogenous; Matu, quasi-viral agent; human; mammary tumour; prion;  
 KM classical virus; slow virus; exogenous MX; p58x; cytoplasmic antigen;  
 KM conservative; Hela cell; twin protein; p54/58N; cell surface; nucleus;  
 KM monoclonal antibody; Mab M75; neoplasm; pre-neoplastic disease; vaccine.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX W09318152-A1.  
 PN  
 XX  
 XX 16-SEP-1993.  
 PD  
 XX  
 XX 08-MAR-1993; 93WO-US002024.  
 PF  
 XX  
 XX 11-MAR-1992; 92CS-00000709.  
 PR  
 XX 21-OCT-1992; 92US-00964589.  
 XX  
 PA (CIBA ) CIBA CORNING DIAGNOSTICS CORP.  
 PA (VIRO-) INST VIROLOGY.  
 XX  
 XX Zavada J, Pastorekova S, Pastorek J;  
 PT  
 DR MPI; 1993-303466/38.  
 XX  
 DR N-PSDB; AAQ8456.  
 XX  
 PT New MN gene and polypeptide(s) - used in diagnosis, prognosis and therapy  
 PT of neoplastic and/or pre-neoplastic disease.  
 PS  
 XX Claim 7; Fig 1; 72pp; English.

CC This sequence is encoded by the intronless MN gene which is a cellular  
 CC gene which is the endogenous component of the Matu agent. Matu is a novel  
 CC quasi-viral agent with rather unusual properties. It is presumably  
 CC derived from a human mammary tumour. In some aspects it resembles  
 CC classical viruses, whereas in other respects it resembles "slow" viruses  
 CC (prions), and in still other aspects it is different from both classes of  
 CC viruses. Matu is a two component system. One part of the complex,  
 CC exogenous MX, is transmissible, and is manifest by a protein, p58x, which  
 CC is a cytoplasmic antigen which reacts with some natural sera, of humans  
 CC and of various animals. The other component, MN, is endogenous to human  
 CC cells. MN is a cellular gene showing very little homology with known DNA  
 CC sequences. It is rather conservative and present as a single copy in the  
 CC chromosomal DNA of various vertebrates. MN is manifest in Hela cells by a  
 CC twin protein p54/58N, that is localised on the cell surface and in the  
 CC nucleus. Immunoblots using a monoclonal antibody reactive with p54/58N  
 CC (Mab M75) reveals two bands at 54 kD and 58 kD. These two bands may  
 CC correspond to one type of protein that differs by glycosylation pattern  
 CC or by how it is processed. The expression of the MN gene is strongly  
 CC correlated with tumorigenicity. MN products can be used in can be used  
 CC in diagnostic and/or prognostic assays for neoplastic and/or pre-  
 CC neoplastic disease. MN polypeptides, produced recombinantly by  
 CC unicellular hosts, can also be used for antibody production and in  
 CC vaccines for inducing protective immunity against neoplastic disease and  
 CC a dampening effect upon tumorigenic activity. (Updated on 25-MAR-2003 to  
 CC correct PN field.)  
 CC  
 XX  
 XX Sequence 429 AA;  
 SQ

Query Match 57.7%; Score 1398.5; DB 2; Length 429;  
 Best Local Similarity 69.1%; Pred. No. 8.7e-102;  
 Matches 300; Conservative 12; Mismatches 83; Indels 39; Gaps 7;

QY 38 QRLPRMOEDSPGCGSSGDDPLGEEDLPSEEDSPREDDPGEEDLPGEEDLP 96  
 DB 1 QRLPRMOEDSPLEEALLGKMTWKARRICVRRIRH-ERRITLERIRYRITYL 59  
 QY 97 EVKPSBEGSLKLEDDPTVEAPDPOEPONNAHDKXGDSQHWYGGDPMPW 150  
 DB 60 KIMPSBEGSLKLEDDPTVEAPDPOEPONNAHDKXGDSQHWYGGDPMPW 119



QY 151 --RVSPACAGRFQSPVDIRPOLAFC-----PALRPLELLGFQLPELRLRNNGH 200  
 Db 120 RAASPRWISAPSSPSARP-----CAPNMSWASSSRSSONCAQCRP----- 162  
 QY 201 SVQLTLPGLGEMALGGRG--YRALQLHFWGAAGRPGSEHTVEGHRPPAEIHVHSTA 256  
 Db 163 --CCATDPASWARDGGSSRAGVPAQLQLHFWGAAGRPGSEHTVEGHRPPAEIHVHSTA 220  
 QY 259 FAVDEALGRPGGLAVLAFLIEGPEENSAVEQLLSRLIEIAEGSETOVGLISALLP 318  
 Db 221 FAVDEALGRPGGLAVLAFLIEGPEENSAVEQLLSRLIEIAEGSETOVGLISALLP 280  
 QY 319 SDPSRYFOYEGSLITTPCAQGVITWVFNQVWMLSAKQHLTSDTLWGPGDSRLQLNFRAT 378  
 Db 281 SDPSRYFOYEGSLITTPCAQGVITWVFNQVWMLSAKQHLTSDTLWGPGDSRLQLNFRAT 340  
 QY 379 QPLNGRVIEASFPAGVDSPPRAAEFYQVNSCLAAQDIALVFGLLFNTSVAFVYQRRQ 438  
 Db 341 QPLNGRVIEASFPAGVDSPPRAAEFYQVNSCLAAQDIALVFGLLFNTSVAFVYQRRQ 400  
 QY 439 HRRGRTKGVSYPRA 452  
 Db 401 HRRGRTKGVSYPRA 414

## RESULT 13

AA53241  
 ID AA53241 standard; protein; 257 AA.

AC AA53241;

DT 16-JUN-2000 (first entry)

DE MN protein carbonic anhydrase domain SEQ ID NO:51.

KM Human; MN protein; MN gene; oncogene; carbonic anhydrase; tumour;

KW oncogenesis; diagnosis; neoplastic disease; cancer; carcinoma;

KX MN/CA IX isoenzyme.

OS Homo sapiens.

PN US6027887-A.

PD 22-FEB-2000.

PF 24-JAN-1997; 97US-00787739.

PR 21-OCT-1992; 92US-00964589.

PR 30-DEC-1993; 93US-00177093.

PR 15-JUN-1994; 94US-00260190.

PR 07-JUN-1995; 95US-00477504.

PR 07-JUN-1995; 95US-00481658.

PR 07-JUN-1995; 95US-00485049.

PR 07-JUN-1995; 95US-00485662.

PR 07-JUN-1995; 95US-00486756.

PR 07-JUN-1995; 95US-00487077.

PA (SISC-) SLOVAK ACAD SCI INST VIROLOGY.

PI Pastorek J, Zavada J, Pastorekova S;

DR WPI, 2000-194827/17.

XX Nucleic acid based assay for diagnosing a wide variety of

PT preneoplastic/neoplastic disease comprises screening for the presence of

CC abnormal MN gene expression in a vertebrate.

CC if abnormal MN gene expression is determined to be present in the  
 CC vertebrate, determining that the vertebrate has a significant risk of  
 CC having preneoplastic/neoplastic disease. The MN gene is an oncogene and  
 CC encodes an MN protein (also referred to as MN/CA IX isoenzyme). The MN  
 CC protein is a tumor associated carbonic anhydrase isoenzyme. The method  
 CC is used for detecting a wide variety of preneoplastic/neoplastic diseases  
 CC in a vertebrate, preferably a human. The disease detected is mammary,  
 CC bladder, renal, urinary tract, ovarian, uterine, cervical, endometrial,  
 CC vaginal, vulval, prostate, liver, lung, skin, thyroid, pancreatic,  
 CC testicular, brain, head and neck, mesodermal, gallbladder, rectal,  
 CC duodenal, jejunal, ileal, gastric, pancreatic duct, liver duct, gastric  
 CC mucosa, gallbladder epithelium, small intestinal mucosa, colorectal  
 CC mucosa, pancreatic duct epithelium or liver duct epithelium  
 CC preneoplastic/neoplastic disease. AA16540 to AA16617 and AA53228 to  
 CC AA53245 represent sequences used in the exemplification of the present  
 CC invention  
 SQ Sequence 257 AA;

Query Match 56.5%; Score 1370; DB 3; Length 257;  
 Best Local Similarity 100.0%; Pred No.7.8e-100; Mismatches 0; Gaps 0;  
 Matches 257; Conservative 0; Indels 0;

QY 135 GDDSHMRVGGDPMPVSPACAGRFQSPVDIRPOLAFCPALRPLELLGFQLPELRLRNNGH 194  
 Db 1 GDDSHMRVGGDPMPVSPACAGRFQSPVDIRPOLAFCPALRPLELLGFQLPELRLRNNGH 60  
 QY 195 LRNGHSVOLTLPPGLEMALGPREYRALQLHFWGAAGRPGSEHTVEGHRPPAEIHVHSTA 254  
 Db 61 LRNGHSVOLTLPPGLEMALGPREYRALQLHFWGAAGRPGSEHTVEGHRPPAEIHVHSTA 120  
 QY 255 LSTAFARVDEALGRPGGLAVLAFLIEGPEENSAVEQLLSRLIEIAEGSETOVGLISALLP 314  
 Db 121 LSTAFARVDEALGRPGGLAVLAFLIEGPEENSAVEQLLSRLIEIAEGSETOVGLISALLP 180  
 QY 315 ALTPSDPSRYFOYEGSLITTPCAQGVITWVFNQVWMLSAKQHLTSDTLWGPGDSRLQLNFRAT 374  
 Db 181 ALTPSDPSRYFOYEGSLITTPCAQGVITWVFNQVWMLSAKQHLTSDTLWGPGDSRLQLNFRAT 240  
 QY 375 FRATQPLNGRVIEASFP 391  
 Db 241 FRATQPLNGRVIEASFP 257

## RESULT 14

AA603018  
 ID AA603018 standard; protein; 257 AA.

AC AA603018;

DT 25-SEP-2000 (first entry)

DE Human MN protein carbonic anhydrase (CA) domain.

KM MN protein; tumour associated cell adhesion molecule; oncoprotein;

KW proteoglycan domain; PG domain; carbonic anhydrase; CA domain;

KX abnormal expression; neoplastic disease; cancer; gene therapy.

OS Homo sapiens.

PN W0200024913-A2.

PD 04-MAY-2000.

PF 22-OCT-1999; 99WO-US024879.

PR 23-OCT-1998; 98US-00177776.

PR 23-OCT-1998; 98US-00178115.

PA (FARB) BAYER CORP.

PI (VIRO-) INST VIROLOGY.

XX Pastorek J, Pastorekova S, Pastorek J;

XX WP1; 2000-350752/30.  
 XX A molecule which specifically binds to a site on MN protein (oncoprotein)  
 PT and prevents adhesion of vertebrate cells to the protein, useful for  
 PT treating preneoplastic or neoplastic diseases such as cancer.  
 XX  
 XX Claim 14; Page 125-126; 154pp; English.  
 XX  
 CC The invention relates to the inhibition of cell adhesion mediated by the  
 CC MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G250  
 CC protein). The MN protein is a tumour-associated adhesion molecule which  
 CC comprises a proteoglycan-like (PG) domain (AAB03017) which contains the  
 CC protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018).  
 CC Abnormal expression of the MN protein is associated with tumorigenicity.  
 CC The invention encompasses molecules (e.g., proteins and peptides) which  
 CC which specifically bind to a site on the MN protein, thereby preventing  
 CC adhesion of vertebrate cells to the protein in a cell adhesion assay. It  
 CC also encompasses MN proteins or MN protein fragments which can be added  
 CC to the extracellular environment to prevent the adhesion of vertebrate  
 CC cells to each other. The invention also relates to the identification of  
 CC the binding site of the MN protein and to a method of identifying a site  
 CC on an MN protein to which cells adhere, comprising testing a series of  
 CC overlapping peptides from the protein in a cell adhesion assay. The  
 CC invention encompasses a vector comprising an expression control sequence  
 CC operatively linked to a nucleic acid encoding the variable domains of a  
 CC MN-specific antibody, where the domains are separated by a flexible  
 CC linker peptide (AAB03035) and the vector inhibits the growth of a  
 CC vertebrate preneoplastic or neoplastic cell that abnormally expresses MN  
 CC protein. The invention also encompasses a vector comprising a nucleic  
 CC acid encoding a cytotoxic protein or peptide operatively linked to the MN  
 CC gene promoter, which inhibits the growth of a vertebrate preneoplastic or  
 CC neoplastic cell. Also claimed is a repressor complex that binds to the MN  
 CC gene promoter (AAB52473). MN proteins and peptides, MN-binding proteins  
 CC and peptides, and expression vectors encoding such proteins and peptides  
 CC are useful for treating patients with preneoplastic or neoplastic diseases  
 CC (e.g., cancers) associated with or characterized by abnormal MN  
 CC expression. Sequences AAB03006, AAB03015-B03022, AAB03029-B03031,  
 CC AAB03036 and AAB03059-B03060 represent specific domains or regions of the  
 CC human MN protein (AAB03005) referred to in the invention  
 XX  
 XX Sequence 257 AA;  
 SQ  
 Query Match 56.5%; Score 1370; DB 3; Length 257;  
 Best Local Similarity 100.0%; Fred. No. 7.8e-100; Indels 0; Gaps 0;  
 Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 135 GDDQSHMRGDDPMPRVSPACAGRFQSPVDIRPOLAFCALRPLELIGFQLPPLPELR 194  
 DB 1 GDDQSHMRGDDPMPRVSPACAGRFQSPVDIRPOLAFCALRPLELIGFQLPPLPELR 60  
 QY 195 LRNGHSHVQLTTPGLENALGGRGYRALQLHLHWGAAGRGSHYVGBRFPPEIHYVHL 254  
 DB 61 LRNGHSHVQLTTPGLENALGGRGYRALQLHLHWGAAGRGSHYVGBRFPPEIHYVHL 120  
 QY 255 LSTAFARVDEALGRGGLAVLAFLFEGBPENSAVEQLSLREIAEGSGSTQVPGDLS 314  
 DB 121 LSTAFARVDEALGRGGLAVLAFLFEGBPENSAVEQLSLREIAEGSGSTQVPGDLS 180  
 QY 315 ALPSPDFSRFYFQYEGSLTTPCAQGVITVFNQIYMSAKOLHTLSLTLMGPGDSRIQLN 374  
 DB 181 ALPSPDFSRFYFQYEGSLTTPCAQGVITVFNQIYMSAKOLHTLSLTLMGPGDSRIQLN 240  
 QY 375 FRATQPLNGRVTEASFP 391  
 DB 241 FRATQPLNGRVTEASFP 257  
 RESULT 15  
 AAR97235  
 ID AAR97235 standard; protein; 256 AA.  
 XX  
 AC AAR97235;

XX 25-MAR-2003 (revised)  
 DT 26-JUL-1996 (first entry)  
 XX  
 DE Multu putative oncogene MN protein carbonic anhydrase domain.  
 XX  
 KW Multu; endogenous; cellular component; MN; Hela cell; diagnosis;  
 KW lymphocytic choriomeningitis virus; LCMV; putative oncogene; treatment;  
 KW neoplastic; pre-neoplastic; disease; antisense therapy; antibody;  
 KW vaccine; vertebrate; immunisation; carbonic anhydrase.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09534650-A2.  
 XX  
 PD 21-DEC-1995.  
 XX  
 PF 15-JUN-1995; 95WC-US007628.  
 XX  
 PR 15-JUN-1994; 94US-00260190.  
 PR 07-JUN-1995; 95US-00477504.  
 PR 07-JUN-1995; 95US-00481658.  
 PR 07-JUN-1995; 95US-00485049.  
 PR 07-JUN-1995; 95US-00485862.  
 PR 07-JUN-1995; 95US-00485863.  
 PR 07-JUN-1995; 95US-00486756.  
 PR 07-JUN-1995; 95US-00487077.  
 XX  
 PA (CIBA ) CIBA CORNING DIAGNOSTICS CORP.  
 PA (VIR-) INST VIROLOGY.  
 XX  
 PI Zavada U; Pastorekova S, Pastorek U;  
 XX  
 DR WP1; 1996-049679/05.  
 XX  
 PT MN gene, protein and nucleic acid fragments - used as primers and probes  
 PT in the detection of MN antigens and antibodies, and in the treatment of  
 PT (pre)neoplastic disease.  
 XX  
 PS Claim 12; -; 102pp; English.  
 XX  
 CC The present sequence is the carbonic anhydrase domain from the protein  
 CC encoded by the full length Multu endogenous cellular component, MN, cDNA  
 CC clone, which was isolated from lymphocytic choriomeningitis virus (LCMV)  
 CC infected Hela cells. Persistent LCMV, the exogenous Multu transmissible  
 CC agent (MX), infection increases the expression level of the MN gene. MN  
 CC is a putative oncogene, and can therefore be used in the development of  
 CC prods. for the diagnosis and treatment of neoplastic (NP), or pre-NP  
 CC diseases. NP diseases can be treated using DNA antisense to MN  
 CC transcribed mRNA, anti-MN protein antibodies can be used for the  
 CC diagnosis NP or pre-NP diseases and a vaccine contg. immunogenic amounts  
 CC of the MN protein can be used to immunise a vertebrate against a NP  
 CC disease associated with MN antigen expression. (Updated on 25-MAR-2003 to  
 CC correct PR field.)  
 XX  
 XX Sequence 256 AA;  
 SQ  
 Query Match 56.3%; Score 1364; DB 2; Length 256;  
 Best Local Similarity 100.0%; Fred. No. 2.3e-99; Indels 0; Gaps 0;  
 Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 136 DDQSHMRGDDPMPRVSPACAGRFQSPVDIRPOLAFCALRPLELIGFQLPPLPELR 195  
 DB 1 DDQSHMRGDDPMPRVSPACAGRFQSPVDIRPOLAFCALRPLELIGFQLPPLPELR 60  
 QY 196 LRNGHSHVQLTTPGLENALGGRGYRALQLHLHWGAAGRGSHYVGBRFPPEIHYVHL 255  
 DB 61 LRNGHSHVQLTTPGLENALGGRGYRALQLHLHWGAAGRGSHYVGBRFPPEIHYVHL 120  
 QY 256 STAFARVDEALGRGGLAVLAFLFEGBPENSAVEQLSLREIAEGSGSTQVPGDLSA 315  
 DB 121 STAFARVDEALGRGGLAVLAFLFEGBPENSAVEQLSLREIAEGSGSTQVPGDLSA 180

Qy	316	LLPDSRRYFQYESSITTPCAQGVITWENQVTWLSAKQHTLISDTLNGPDSRIQLNF	375
Db	181	LLPDSRRYFQYESSITTPCAQGVITWENQVTWLSAKQHTLISDTLNGPDSRIQLNF	240
Qy	376	RATQPLNGRVIASFP	391
Db	241	RATQPLNGRVIASFP	256

Search completed: March 5, 2004, 09:04:10  
Job time : 61 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: March 5, 2004, 09:02:10 ; Search time 20 Seconds

(without alignments)  
2207.596 Million cell updates/sec

Title: US-09-967-237-2

Perfect score: 2424

Sequence: 1 MAPLCPSPMLPILIPAPAPG.....RRGTGVSYPAPAEVETGA 459

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

PIR\_78:\*  
1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	* Match Length	DB	ID	Description
1	2424	100.0	459	2	I38013
2	526	21.7	307	2	A23993
3	511	21.1	319	2	S71877
4	492	20.3	308	1	CRH06
5	425	17.5	303	2	S12867
6	412	17.0	290	2	UN0576
7	409	16.9	264	1	CRH07
8	402.5	16.6	260	2	T08463
9	391	16.1	259	1	CRB02
10	380.5	15.7	260	1	CRMS2
11	379.5	15.7	260	1	CRH02
12	379	15.6	259	1	CRSH2
13	374.5	15.4	259	1	CRB2
14	368.5	15.2	260	2	UH0527
15	366.5	15.1	260	1	CRMQ1R
16	364.5	15.0	260	2	JC2580
17	363	15.0	261	2	A26344
18	362.5	15.0	261	2	UN0836
19	359.5	14.8	260	2	I52551
20	359.5	14.8	261	1	CRH01
21	359.5	14.8	261	1	UN0835
22	354	14.6	260	2	CRH01D
23	354	14.6	260	2	A43641
24	347	14.3	309	2	I51900
25	343.5	14.2	260	1	CRH03
26	342	14.1	304	2	I59261
27	338.5	14.0	259	2	A22612
28	333	13.7	334	2	T16772
29	328.5	13.6	305	1	CRH05

30	326.5	13.5	1442	1	B46148	protein-tyrosine-p
31	325	13.4	298	2	S12579	carbonate dehydrat
32	323	13.3	1445	1	A46148	protein-tyrosine-p
33	322	13.3	246	2	T16575	hypothetical prote
34	321	13.2	235	2	A23962	carbonate dehydrat
35	304.5	12.6	2314	1	A46151	protein-tyrosine-p
36	303	12.5	1422	2	T42636	protein-tyrosine-p
37	301	12.4	306	2	A46993	protein-tyrosine-p
38	288.5	11.9	312	1	CRH04	carbonate dehydrat
39	288.5	11.3	319	2	T15137	hypothetical prote
40	270.5	11.3	275	2	H83694	carbonic anhydrase
41	270.5	11.2	324	2	T14196	hypothetical prote
42	268.5	11.1	259	2	T23849	hypothetical prote
43	264.5	10.9	328	2	UB0375	carbonic anhydrase
44	263.5	10.9	239	2	F82479	carbonic anhydrase
45	261.5	10.8	365	2	F88449	protein F54D8.4 [1

## ALIGNMENTS

RESULT 1  
I38013  
p54/58N - human  
C:Species: Homo sapiens (man)  
C:Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 05-Nov-1999  
C:Accession: I38013  
R:Pastorek, U.; Pastorekova, S.; Callebaut, I.; Morron, U.; Zelnik, V.; Opavsky, R.; Oncogene 9, 2877-2888, 1994  
A:Title: Cloning and characterization of MN, a human tumor-associated protein with a A:reference number: I38013; MUID:94366734; PMID:8084592  
A:Accession: I38013  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-459 <RES>  
A:Cross-references: EMBL:X66839; NID:G1000701; FIDN:CAA47315.1; PID:G1000702  
C:Gene: MatuN  
C:Superfamily: carbonic anhydrase homology  
F:141-390/Domain: carbonic anhydrase homology <CNH>

Query Match	100.0%	Score 2424;	DB 2;	Length 459;
Best Local Similarity	100.0%	Pred. No. 4.1e-147;		
Matches 459;	Conservative	0;	Mismatches 0;	Indels 0;
QY	1	MAPLCPSPMLPILIPAPAPGLTVQLLSLLIMVHPORLPROMEDSPLGGSSGSDPL	60	
DB	1	MAPLCPSPMLPILIPAPAPGLTVQLLSLLIMVHPORLPROMEDSPLGGSSGSDPL	60	
QY	61	GSEDLPSEEDSFRSEDPGSEEDLPGEEDLPGEEDLPVKKPKSEEGSLKLEDLPTVAPG	120	
DB	61	GSEDLPSEEDSFRSEDPGSEEDLPGEEDLPGEEDLPVKKPKSEEGSLKLEDLPTVAPG	120	
QY	121	DQEPONNAHDKEDDQSHRWYGGDPPMPVSPACAGRPSPYDIPOLAAFCPALRPL	180	
DB	121	DQEPONNAHDKEDDQSHRWYGGDPPMPVSPACAGRPSPYDIPOLAAFCPALRPL	180	
QY	181	ELLGQLPPLPELRIRNNGSHVQLTTPGLEMALGPGREYRALQLHFWAAGPQSEHT	240	
DB	181	ELLGQLPPLPELRIRNNGSHVQLTTPGLEMALGPGREYRALQLHFWAAGPQSEHT	240	
QY	241	VEGHFPPEIHVHLSFAFVDEALRPGSLATLAFLEGGPENSAYEQLLSRLEIA	300	
DB	241	VEGHFPPEIHVHLSFAFVDEALRPGSLATLAFLEGGPENSAYEQLLSRLEIA	300	
QY	301	EEGSETQVPGDLISALPSDFSRFYQEGSLTPPCAGQVITVFQNTWLSAKQLTTLG	360	
DB	301	EEGSETQVPGDLISALPSDFSRFYQEGSLTPPCAGQVITVFQNTWLSAKQLTTLG	360	
QY	361	DTLWSPGSRQLNPRATOPNGRVIASPPAGVDSSPRAAPVQNSCAAGDITLAVF	420	
DB	361	DTLWSPGSRQLNPRATOPNGRVIASPPAGVDSSPRAAPVQNSCAAGDITLAVF	420	

QY 421 GILFATSVAFVQMRQRGRGTGGVSRPAEVAETGA 459  
 DB 421 GILFATSVAFVQMRQRGRGTGGVSRPAEVAETGA 459

## RESULT 2

A29993  
 C:Enzyme: carbonic dehydratase (EC 4.2.1.1) VI - sheep  
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
 C>Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 23-May-1997  
 C/Accession: A29993  
 R:Bernley, R.T.; Wright, R.D.; Coghlan, J.P.  
 Biochemistry 27, 2815-2820, 1988  
 A>Title: Complete amino acid sequence of ovine salivary carbonic anhydrase.  
 A:Reference number: A29993; PMID:88294021; PMID:3135834  
 A:Accession: A29993  
 A:Molecule type: protein  
 A:Residues: 1-307 <PER>

A:Experimental source: parotid gland  
 A>Note: 63-Mer and 297-11e were also found  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C/Keywords: carbon-oxygen lyase; hydro-lyase; zinc  
 F:6-261/Domain: carbonic anhydrase homology <CAH>

Query Match 21.7%; Score 526; DB 2; Length 307;  
 Best Local Similarity 43.2%; Pred. No. 1.9e-26;  
 Matches 111; Conservative 40; Mismatches 98; Indels 8; Gaps 4;

QY 141 WRG---GDPMRPRSPACGRFQSPVDIRPOLAFCPALRPLELGGFQLPPLRLR 196  
 DB 6 WYSEGMDEAMHPELEPKCGGRSPDLOMKKYQVNSRANLNGYLMH-GEFPT 64  
 QY 197 NNGHSVQLTPPGLEMLAGREYRALQLHLHWGAAGR--PGSEHTVEGHRPAEIHVH 254  
 DB 65 NNGHTVQLSPSTMTSDGTQYLAKQMFHWGASSEISGSEHTVDGKRYIEIHVH 124  
 QY 255 LSTAFARVDEALGRPGGLAVLAFL-EGPEENSAYEQLSRLEIAEGSETOVPGDLI 313  
 DB 125 YNSKINSYEAKQEPDGLAVLAIVEKDYENAYYSFISLHEDIRYAGQSTVLRGLDI 184  
 QY 314 SALLPSDFSRFYQEGSLTTPPCAGVIWVFNQTMLSAKQLTSLDTLWGPDSRLQL 373  
 DB 185 EDLPGDLIRYYYSGLTTPPCTEVHMFVADTVKLSKQVEKLENSLHNQNTION 244  
 QY 374 NFRATQPLNGRIEASF 390  
 DB 245 DFRITQPLNHRVVEANF 261

## RESULT 3

S71877  
 C:Enzyme: carbonic dehydratase (EC 4.2.1.1) isoform VI precursor - bovine  
 N/Alternate names: carbonic anhydrase VI  
 C/Species: Bos primigenius taurus (cattle)  
 C>Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 20-Jun-2000  
 C/Accession: S71877; S03862  
 R:Jiang, W.; Wolbach, J.T.; Gupta, D.  
 Biochem. J. 318, 291-296, 1996  
 A>Title: Sequence of bovine carbonic anhydrase VI: potential recognition sites for N-ace  
 A:Reference number: S71877; PMID:96558528; PMID:8761494  
 A:Accession: S71877  
 A:Molecule type: mRNA  
 A:Residues: 1-319 <UIA>

A:Cross-references: EMBL:X96503; NID:g1526571; PIDN:CAA65357.1; PID:g1526572  
 A:Experimental source: submandibular gland  
 R:Bernley, R.T.; Darling, P.; Aldred, P.; Wright, R.D.; Coghlan, J.P.  
 Biochem. J. 259, 91-96, 1989  
 A>Title: Tissue and species distribution of the secreted carbonic anhydrase isoenzyme.  
 A:Reference number: S03862; PMID:89246311; PMID:2497732  
 A:Accession: S03862  
 A:Molecule type: protein  
 A:Residues: 15,'S',17-36,'X',38-39 <FER>  
 C/Genetics:

A:Gene: CAH6  
 C:Function:  
 A:Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C/Keywords: carbon-oxygen lyase; glycoprotein; hydro-lyase; metalloprotein; zinc  
 F:1-14/Domain: signal sequence #status predicted <SIG>  
 F:15-319/Product: carbonic dehydratase isoform VI #status experimental <MAT>  
 F:18-273/Domain: carbonic anhydrase homology <CAH>  
 F:37-219/Disulfide bonds: #status predicted  
 F:62,251/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:106,108,133/Binding site: zinc (His) #status predicted

Query Match 21.1%; Score 511; DB 2; Length 319;  
 Best Local Similarity 42.0%; Pred. No. 1.8e-25;  
 Matches 108; Conservative 40; Mismatches 97; Indels 12; Gaps 4;

QY 137 DQSHMRVGGDPMPRVPSPACGRFQSPVDIRPOLAFCPALRPLELGGFQLPPLRLR 196  
 DB 26 DEKHRL-----QYPCGQGTQSPIDLMKKVRYPSPRLANLTGYGLRQ-GEFPT 76  
 QY 197 NNGHSVQLTPPGLEMLAGREYRALQLHLHWGAAGR--AAGRPGSEHTVEGHRPAEIHVH 254  
 DB 77 NNGHTVQLSPSTMTSDGTQYLAKQMFHWGASSEISGSEHTVDGKRYIEIHVH 136  
 QY 255 LSTAFARVDEALGRPGGLAVLAFL-EGPEENSAYEQLSRLEIAEGSETOVPGDLI 313  
 DB 137 YNSKINSYEAKQEPDGLAVLAIVEKDYENAYYSNFISLHEDIRYAGQSTVLRGLDI 196  
 QY 314 SALLPSDFSRFYQEGSLTTPPCAGVIWVFNQTMLSAKQLTSLDTLWGPDSRLQL 373  
 DB 197 QDMPLGDLIRYYYSGLTTPPCTEVHMFVADTVKLSKQVEKLENSLHNQNTION 256  
 QY 374 NFRATQPLNGRIEASF 390  
 DB 257 NYRSTQPLNHRVVEANF 273

## RESULT 4

C8R06  
 C:Enzyme: carbonic dehydratase (EC 4.2.1.1) VI precursor - human  
 N/Alternate names: carbonic anhydrase VI; salivary carbonic anhydrase; secreted carb  
 C/Species: Homo sapiens (man)  
 C>Date: 31-May-1991 #sequence\_revision 05-May-1995 #text\_change 02-Sep-1997  
 C/Accession: A37917  
 R:Aldred, P.; Fu, P.; Barrett, G.; Penschow, J.D.; Wright, R.D.; Coghlan, J.P.; Fern  
 Biochemistry 30, 569-575, 1991  
 A>Title: Human secreted carbonic anhydrase: cDNA cloning, nucleotide sequence, and h  
 A:Reference number: A37917; PMID:9105141; PMID:1899030  
 A:Accession: A37917  
 A:Molecule type: mRNA  
 A:Residues: 1-308 <AUD>

A:Cross-references: GB:M57892; GB:J05305  
 A:Experimental source: salivary gland  
 A>Note: The authors translated the codon GAG for residue 248 as Gln  
 C/Genetics:

A:Gene: GDB: CAG  
 A:Cross-references: GDB:125350; OMIM:114780  
 A:Map position: 1936-1936  
 C:Function:

A:Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide  
 A>Note: This form is expressed in salivary, parotid, and submandibular glands; it is  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C/Keywords: carbon-oxygen lyase; glycoprotein; hydro-lyase; metalloprotein; pyroglyut  
 F:1-17/Domain: signal sequence #status predicted <SIG>  
 F:18-308/Product: carbonic dehydratase VI #status predicted <MAT>  
 F:23-278/Domain: carbonic anhydrase homology <CAH>  
 F:118/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predi  
 F:442-224/Disulfide bonds: #status predicted  
 F:67,256/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:111,113,138/Binding site: zinc (His) #status predicted

Query Match 20.3%; Score 492; DB 1; Length 308;  
 Best Local Similarity 39.6%; Pred. No. 2.8e-24;

Matches 103; Conservative 44; Mismatches 105; Indels 8; Gaps 4;

QY 139 SHRRY-----GDPWRVSDACAGRFQSPVDIRPOLAFCPALRPLELIGFOLPELR 194  
 Db 21 SDMTYEGALDEAHMQRHPACGGGQSPINLQRTKVRVYNPLKLMNTGTYE--IQAGFFP 79

QY 195 LRNGHSVQTLTPPGEMALGPGRERFALQJLHFWAGAGR--PGEHTEGHRRPAEIHV 252  
 Db 80 MANNHHTVQIGLPTSMRTMYADQIVYIAQQMFHMGASSEISGSEHTVDGIRHYEIH 139

QY 253 VHSFAFVDAALGRPGGIATVIAFL-EGEENSAYEQLLSRLEEAEGSETOVPG 311  
 Db 140 VHNYSKRYKTDIQDAPDGLAVLAFAVEYKXNPENTYVSNFISHLANIKYPGQRTLTGL 199

QY 312 DISALLPSDFSRFYEGESLTPPGAGQVIMTVFQGTWLSAKQJHLSDTLMDPGSR 371  
 Db 200 DVQDMLPRNQHYYTHGSLTTPCTENVMFVLADFKLSRTQVWKLNSLDHRNKTI 259

QY 372 QLNFRATOPUNGIVIASFP 391  
 Db 260 HNDYRTTOPLNRHVESNFP 279

## RESULT 5

S12867  
 carbonate dehydratase-related protein - mouse  
 N/Alternate names: carbonic anhydrase-related protein  
 C/Species: Mus musculus (house mouse)  
 C/Date: 18-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999  
 C/Accession: S12867  
 R/Kato, K.  
 FEBS Lett. 271, 137-140, 1990

A/Title: Sequence of a novel carbonic anhydrase-related polypeptide and its exclusive pr  
 A/Reference number: S12867; MUID:91032050; PMID:2121526  
 A/Accession: S12867  
 A/Molecule type: mRNA  
 A/Residues: 1-303 <KAT>  
 A/Cross-references: EMBL:X61397; NID:950286; PIDN:CAA3668.1; PID:950287  
 C/Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 F/30-290/Domain: carbonic anhydrase homology <CAH>

Query Match 17.5%; Score 425; DB 2; Length 303;  
 Best Local Similarity 32.9%; Pred. No. 5e-20;  
 Matches 104; Conservative 61; Mismatches 119; Indels 32; Gaps 10;

QY 110 LEDLPTVE-APGDPOEPQNNARDEGDQSHWRYGDPMPRVSPACAGRFQSPVDIRP 168  
 Db 1 MADLSFIEDVAFAPEKEDEDEEEEGVE--WGYEKGVEGLVFPDANGRYQSPINLNS 57

QY 169 QLAAPCALRPLELIGFOLP---LPELRLRNNGHSVQTLTPGLEMALGP---GREYR 221  
 Db 58 REARVDP-----LDVRLSPYVVCRCCEVTNDGHTIQVILKSKSVLSGGLPQGGFE 112

QY 222 ALQHLHWGAGRPGESEHTVEGHRPAEIHVHL-STAFAVDAALGRPGGLAVLAFL 280  
 Db 113 LVEVFHWGRENQORSEHTVFKAPFMELHILHNNSTLFGSIDAVGKPGHAIIALFVQ 172

QY 281 EGPENSAYEQLLSRLEEAEGSETOVPGDISALLPSDSR-YFYEGSLTPPGAG 339  
 Db 173 IG-KHVGKAVTELLQDIQYKSKTIPCNFPTLLPDLRLRYWYBSGLTTPGSE 231

QY 340 VITVFNQTVMSAKQJ-----HTLSDTLWPGDSRQLNFRATOPUNGIVIASFP 391  
 Db 232 VIMILFRYPLTISQIQIEFRRLRTHVKGALVEGCGIIGDNRFPQPLSDRVITRAAF 290

QY 392 AGVDSPPRAEYVQIN 407  
 Db 291 -----SSQDRREGTTLH 302

RESULT 6  
 JN0576  
 carbonate dehydratase (EC 4.2.1.1) VIII - human

N/Alternate names: carbonic anhydrase-related protein (CARP)  
 C/Species: Homo sapiens (man)  
 C/Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 21-Jul-2000  
 C/Accession: JN0576  
 R/Skaggs, L.N.; Berghem, N.C.H.; Venta, P.J.; Tashian, R.E.  
 Gene 126, 291-292, 1993

A/Title: The deduced amino acid sequence of human carbonic anhydrase-related protein  
 A/Reference number: JN0576; MUID:93246262; PMID:8482548

A/Accession: JN0576  
 A/Molecule type: mRNA  
 A/Residues: 1-290 <SKA>  
 A/Cross-references: GB:L04656; NID:9179937; PIDN:AAA3653.2; PID:95069431  
 C/Genetic:

A/Genes: GDB:CA9; CALS; CARP  
 A/Cross-references: GDB:141839; OMIM:114815  
 A/Map position: 8pter-8qter  
 C/Function:  
 A/Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxi  
 A/Note: this protein has not been demonstrated to be enzymatically active  
 C/Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C/Keywords: carbon-oxygen lyase; hydro-lyase; metalloprotein; zinc  
 F/2-261/Product: carbonate dehydratase VIII #status predicted <WAT>  
 F/29-289/Domain: carbonic anhydrase homology <CAH>

Query Match 17.0%; Score 412; DB 2; Length 290;  
 Best Local Similarity 32.9%; Pred. No. 3.2e-19;  
 Matches 98; Conservative 57; Mismatches 117; Indels 26; Gaps 8;

QY 110 LEDLPTVAAPGDPOEPQNNARDEGDQSHWRYGDPMPRVSPACAGRFQSPVDIRP 169  
 Db 1 MADLSFIEDVAFAPEKEDEDEEEEGVE--WGYEKGVEGLVFPDANGRYQSPINLNS 57

QY 170 LAAPCALRPLELIGFOLP---LPELRLRNNGHSVQTLTPGLEMALGP---GREYR 222  
 Db 58 REARVDP-----LDVRLSPYVVCRCCEVTNDGHTIQVILKSKSVLSGGLPQGGFE 112

QY 223 LQHLHWGAGRPGESEHTVEGHRPAEIHVHL-STAFAVDAALGRPGGLAVLAFL 281  
 Db 113 LVEVFHWGRENQORSEHTVFKAPFMELHILHNNSTLFGSIDAVGKPGHAIIALFVQ 172

QY 282 EGPENSAYEQLLSRLEEAEGSETOVPGDISALLPSDSR-YFYEGSLTPPGAG 340  
 Db 173 IG-KHVGKAVTELLQDIQYKSKTIPCNFPTLLPDLRLRYWYBSGLTTPGSE 231

QY 341 VITVFNQTVMSAKQJ-----HTLSDTLWPGDSRQLNFRATOPUNGIVIASF 390  
 Db 232 VIMILFRYPLTISQIQIEFRRLRTHVKGALVEGCGIIGDNRFPQPLSDRVITRAAF 289

## RESULT 7

CRH07  
 carbonate dehydratase (EC 4.2.1.1) VII - human  
 N/Alternate names: carbonic anhydrase VII

C/Species: Homo sapiens (man)  
 C/Date: 10-Feb-1995 #sequence\_revision 05-May-1995 #text\_change 18-Jun-1999  
 C/Accession: A55272

R/Montgomery, J.C.; Venta, P.J.; Eddy, R.L.; Fukushima, Y.S.; Shown, T.B.; Tashian, R.  
 Genomics 11, 835-848, 1991  
 A/Title: Characterization of the human gene for a newly discovered carbonic anhydrase  
 A/Reference number: A55272; MUID:92147127; PMID:1783392

A/Accession: A55272  
 A/Molecule type: DNA  
 A/Residues: 1-264 <MON>  
 A/Cross-references: GB:W66423; NID:9179964; PIDN:AAA5193.1; PID:9179967  
 A/Note: sequence extracted from NCBI Backbone (NCBI:80199, NCBI:80201, NCBI:80205,  
 C/Genetic:

A/Genes: GDB:CA7  
 A/Cross-references: GDB:119741; OMIM:114770  
 A/Map position: 16q22.1-16q22.1  
 A/Introns: 14/1; 80/1; 119/3; 151/3; 172/3; 224/3  
 C/Function:  
 A/Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide  
 A/Note: this form is expressed in salivary gland and other tissues

C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C:Keywords: carbon-oxygen lyase; hydro-lyase; metalloprotein; zinc  
 F/4-264/Product: carbonate dehydratase VII #status predicted <MAT>  
 F/7-262/Domain: carbonic anhydrase homology <CAH>  
 F/56/58/121/Binding site: zinc (His) #status predicted

Query Match 16.9%; Score 409; DB 1; Length 264;  
 Best Local Similarity 37.5%; Pred. No. 4,4e-19;

Matches 101; Conservative 41; Mismatches 99; Indels 28; Gaps 11;

QY 141 WRYG---GDPWPRVSPACAGFQSPVDIRPOLAFCPALPPELLGQPLPPELRN 197  
 DB 7 WYGGDDGSHMKLYPIAQGRQSPINISSQAVYSPSLQPLEL---SYEACMSLSTW 63  
 QY 198 NCHSVOL-----TLPRG--LEMALGRREYALQHLHMGAAAGPSSHEHYEGRFP 247  
 DB 64 NCHSVQVDPNDSDDTIVTGGPLE---GP---YIKQPHFHWGKKHDVSGSEHTVDGKSP 117  
 QY 248 AEIHYVHLST-AFAVDEALGRPGGLAVLAFLIEGEPENSAYEQLRLLEIAEGSEST 306  
 DB 118 SELHYVHWMAKKYSTFGAASAPDGLAVGVFLETG-DEHPSMRLTDLAYVVRFGKTXA 176  
 QY 307 QYRGDISALLPSDSRYFOYEGSLTTPPCAGQVITWYFNQVWLSKQHLTSLTM-G 365  
 DB 177 QSCFNPCKLPPAS-RHYWTYPSLTPPLESEVTVIWRPCLISERQKGRSLLFTS 235  
 QY 366 PGDSRLQL--NFRATQPLNGRVIEASFP 392  
 DB 236 EDDERIHVWNNRPPQPLKGRVYKASFP 264

## RESULT 8

T08463  
 Carbonate dehydratase (EC 4.2.1.1) - zebra fish  
 C:Species: Brachydanio rerio (zebra fish)  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999  
 C:Accession: T08463  
 R:Peterson, R.E.; Tu, C.; Linser, P.J.  
 J. Mol. Evol. 44, 432-439, 1997  
 A>Title: Isolation and characterization of a carbonic anhydrase homologue from the zebra  
 A:Reference number: Z16422; MUID:97250269; PMID:9089083  
 A:Accession: T08463  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-260 <PRT>  
 A:Cross-references: EMBL:U55177; NID:92576334; PID:92576335  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C:Keywords: carbon-oxygen lyase; hydro-lyase; zinc  
 F/5-259/Domain: carbonic anhydrase homology <CAH>

Query Match 16.6%; Score 402.5; DB 2; Length 260;  
 Best Local Similarity 35.8%; Pred. No. 1.1e-18;  
 Matches 95; Conservative 37; Mismatches 108; Indels 25; Gaps 7;

QY 141 WRYG---GDPWPRVSPACAGFQSPVDIRPOLAFCPALPPELLGQPLPPELRN 197  
 DB 5 WYGGADGPESMAESFFPIANGPROSPIDIVTQHPDSIKHKL---KYDPATKSLIN 61  
 QY 198 NCHSVOL-----TLPRG--LEMALGRREYALQHLHMGAAAGPSSHEHYEGRFP 248  
 DB 62 NCHSVQVDPNDSDDTIVTGGPLE---GP---YIKQPHFHWGKKHDVSGSEHTVDGKSP 117  
 QY 249 EIHVHLSTAFARVDEALGRPGGLAVLAFLIEGEPENSAYEQLRLLEIAEGSEST 308  
 DB 117 EHLHYVHWMAKKYSTFGAASAPDGLAVGVFLETG-AANPLQLKVLADLIDIKSKGQTF 175  
 QY 309 PGDISALLPSDSRYFOYEGSLTTPPCAGQVITWYFNQVWLSKQHLTSLTM---G 365  
 DB 176 ANFPCKTLPPASL-YWYBGSLLTPPLLSVITVILKEPISVSQAKMFKRSLPSSG 234  
 QY 366 PGDSRLQLNFRATQPLNGRVIEASFP 390  
 DB 235 ETPCCMVNRRPPQPLKGRKVRASF 259

RESULT 9  
 CRB02  
 Carbonate dehydratase (EC 4.2.1.1) II - bovine (tentative sequence)

N:Alternate names: carbonic anhydrase II  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 31-Mar-1991 #sequence\_revision 31-Mar-1981 #text\_change 31-Mar-2000  
 C:Accession: A01144

R:Scakly, M.; Limozin, N.; Filippi-Foveau, D.; Guilian, J.M.; Laurent-Tabusse, G.  
 Biochimie 58, 1071-1082, 1976  
 A>Title: Structure primaire de l'anhydrase carbonique erythrocytaire bovine CI. II.

A:Reference number: A90672; MUID:77065798; PMID:826282

A:Accession: A01144  
 A:Molecule type: protein  
 A:Residues: 1-259 <SCI>  
 R:Guilian, J.M.; Limozin, N.; Mallet, B.; Di Costanzo, J.; Charrel, M.

Biochimie 59, 293-302, 1977  
 A>Title: Independance genetique de deux formes de l'anhydrase carbonique erythrocytaire  
 A:Reference number: A90672; MUID:77242599; PMID:19093

A:Contents: annotation  
 A>Note: One minor and two major forms were isolated chromatographically. One of the  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc  
 F/4-258/Domain: carbonic anhydrase homology <CAH>  
 F/1/Modified site: acetylated amino end (Ser) #status experimental  
 F/93/95/118/Binding site: zinc (His) #status predicted

Query Match 16.1%; Score 391; DB 1; Length 259;  
 Best Local Similarity 35.5%; Pred. No. 6.1e-18;  
 Matches 93; Conservative 43; Mismatches 110; Indels 16; Gaps 8;

QY 140 HMYG---GDPWPRVSPACAGFQSPVDIRPOLAFCPALPPELLGQPLPPELRN 196  
 DB 3 HMYGKKBGZDZHHKDFPIANGPROSPVIDTAVVQDPLAPLVL---YEATSRNRY 59  
 QY 197 NCHSVOLTLPRGEMAL---GP-GREYALQHLHMGAAAGPSSHEHYEGRFAELHV 252  
 DB 60 NCHSVFVEXYDSDQAVLKDGLTGTLYVDFHFGWSSBQSGSEHTVDKRYAAEHL 119  
 QY 253 VHLSTARVDEALGRPGGLAVLAFLIEGEPENSAYEQLRLLEIAEGSESTVPGD 312  
 DB 120 VHWMTXGDBGTAQCPDGLAVGVFLKYG-DANPLQLKVLADLIDIKSKGSTFPPND 178  
 QY 313 ISALLPSDSRYFOYEGSLTTPPCAGQVITWYFNQVWLSKQHLTSLTM-G 365  
 DB 179 PGSLTPPVLD-YWYBGSLLTPPLLSVITVILKEPISVSQAKMFKRSLPSSG 237  
 QY 370 RLQNFATQPLNGRVIEASFP 391  
 DB 238 LMLANRRPPQPLKGRKVR-GFP 256

## RESULT 10

CRMS2  
 Carbonate dehydratase (EC 4.2.1.1) II - mouse

N:Alternate names: carbonic anhydrase II  
 C:Species: Mus musculus (house mouse)  
 C>Date: 25-Feb-1985 #sequence\_revision 06-Feb-1995 #text\_change 18-Jun-1995  
 C:Accession: A23900; B23202; A01143; A20535; I51549

R:Ventra, P.J.; Montgomey, J.C.; Hewett-Emmett, D.; Wiebauer, K.; Tashian, R.E.  
 J. Biol. Chem. 260, 12130-12135, 1985  
 A>Title: Structure and exon to protein domain relationships of the mouse carbonic an

A:Reference number: A23900; MUID:86008276; PMID:2995362

A:Accession: A23900  
 A:Molecule type: DNA

A:Residues: 1-260 <VEN>  
 A:Experimental source: strain YBR

R:Ventra, P.J.; Montgomey, C.; Hewett-Emmett, D.; Tashian, R.E.  
 Biochim. Biophys. Acta 826, 195-201, 1985

A>Title: Comparison of the 5' regions of human and mouse carbonic anhydrase II genes  
 A:Reference number: A90655; MUID:86077780; PMID:3000449  
 A:Accession: B23202



A/Molecule type: DNA  
 A/Residues: 1-77 <VE2>  
 A/Note: the authors translated the codon CAG for residue 39 as His  
 R/Curtis, P.J.; Wilbers, E.; Demuth, D.; Wett, R.; Venter, P.J.; Tashian, R.E.  
 Gene 25, 325-332, 1993  
 A/Title: The nucleotide sequence and derived amino acid sequence of cDNA coding for mouse  
 A/Reference number: A01143; MUID:84109569; PMID:6420240  
 A/Accession: A01143  
 A/Molecule type: mRNA  
 A/Residues: 2-38, 'H', 40-260 <CUR>  
 A/Cross-references: GB:K00812; GB:M11330; NID:G192333; PIDDN:AAA7356.1; PID:G  
 A/Note: Initiator Met not shown  
 R/Curtis, P.J.  
 J. Biol. Chem. 258, 4459-4463, 1983  
 A/Title: Cloning of mouse carbonic anhydrase mRNA and its induction in mouse erythrocyte  
 A/Reference number: A20539; MUID:83161023; PMID:6187736  
 A/Accession: A20539  
 A/Molecule type: mRNA  
 A/Residues: 155-178, 214-240 <CU2>  
 R/Venter, P.J.; Montgomery, J.C.; Wiebauer, K.; Hewett-Emmett, D.; Tashian, R.E.  
 Ann. N. Y. Acad. Sci. 429, 309-323, 1984  
 A/Title: Organization of the mouse and human carbonic anhydrase II genes.  
 A/Reference number: I51949; MUID:84555152; PMID:6331255  
 A/Accession: I51949  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 241-260 <RES>  
 A/Cross-references: GB:M25944; NID:G199078; PIDDN:AAA39505.1; PID:G199079  
 A/Accession: M25944  
 A/Status: Car-2  
 C/Genetics:  
 A/Introns: 12/1; 78/1; 117/3; 144/1; 169/3; 221/3  
 C/Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C/Keywords: carbon-oxygen lyase; hydro-lyase; zinc  
 F/5-285/Domain: carbonic anhydrase homology <CAH>  
 F/4, 96, 119/Binding site: zinc (His) #status predicted

Query Match 15.7%; Score 380.5; DB 1; Length 260;  
 Best Local Similarity 35.6%; Pred. No. 2,8e-17;  
 Matches 93; Conservative 41; Mismatches 112; Indels 15; Gaps 6;

QY 140 HRRY--GGDPWPRVSPACAGRFSPVDIRPOLAFAFCPALRPLELGFQPLPELRRLR 196  
 DB 4 HNGYSKHNKPNENHNDPFLANDRQSPVDIDTAQHDPALQPL--LISYD--KAASKSIY 60  
 QY 197 NNGHSVOLTLPPGLEMAL--GP--GREYALQHLHMGACRPGSEETVGHPPAEIHY 252  
 DB 61 NNGHSFNVFDDSCONAVLKGRPLSDSYRLQFHFHMOSSDQGSSEITVKKKYALEHL 120  
 QY 253 VHLSTAFARVDEALRPGGLAVLAFLKEGPEENSAYEQILSLLEIAEGSETOVPGLD 312  
 DB 121 VMNNTKYGDFGKAVQOPDGLAVLGYFLKIGP--ASQGGKVEALHSIKTKGKAAPANFD 179  
 QY 313 ISALLPSDSRYFOYEGSLTTPPCAGGYIVFVQOTWLSAKOL--HTLSDTLMPGDS 369  
 DB 180 PGLSLPGNLD--YMYPSGLTTPPLELCTVWLVLEPITVSSQMSHRTFLNVEEDGAE 238  
 QY 370 RLQNLFRATQPLNGEIVASFP 390  
 DB 239 AMVDNMRPAQPLNKKIKASF 259

RESULT 11  
 CRU2  
 carbonate dehydratase (EC 4.2.1.1) II [validated] - human  
 N/A/Alternate names: carbonic anhydrase II; hepatic carbonic anhydrase  
 C/Species: Homo sapiens (man)  
 C/Date: 07-May-1981 #sequence revision 05-May-1995 #text change 15-Sep-2000  
 C/Accession: A27175; A23202; A92147; I37214; I51663; I51877; A01141  
 R/Murakami, H.; Marelich, G.P.; Grubb, J.H.; Kyle, J.W.; Sly, W.S.  
 Genomics 1, 159-166, 1987  
 A/Title: Cloning, expression, and sequence homologies of cDNA for human carbonic anhydrase  
 A/Reference number: A27175; MUID:88085190; PMID:3121456  
 A/Accession: A27175

A/Molecule type: mRNA  
 A/Residues: 1-260 <MR>  
 A/Cross-references: GB:J03037; NID:G197771; PIDDN:AAA51908.1; PID:G197772  
 R/Venter, P.J.; Montgomery, C.; Hewett-Emmett, D.; Tashian, R.E.  
 Biochim. Biophys. Acta 826, 195-201, 1985  
 A/Title: Comparison of the 5' regions of human and mouse carbonic anhydrase II genes  
 A/Reference number: A90655; MUID:86077780; PMID:3000449  
 A/Accession: A90655  
 A/Molecule type: DNA  
 A/Residues: A23202  
 A/Cross-references: GB:X03251; GB:M18100; GB:M77181; NID:G197978; PIDDN:AAA51909.1; PI  
 R/Henderson, L.E.; Henriksson, D.; Nyman, P.O.  
 J. Biol. Chem. 251, 5457-5463, 1976  
 A/Title: The primary structure of human carbonic anhydrase C.  
 A/Reference number: A92194; MUID:7706079; PMID:823150  
 A/Accession: A92194  
 A/Molecule type: protein  
 A/Residues: 2-260 <HEN>  
 R/Ult, K.T.D.; Deutsch, H.F.  
 J. Biol. Chem. 249, 2329-2337, 1974  
 A/Title: Human carbonic anhydrases. XII. The complete primary structure of the C iso  
 A/Reference number: A92147; MUID:74143468; PMID:4207120  
 A/Accession: A92147  
 A/Molecule type: protein  
 A/Residues: 2-260 <LIN>  
 R/Montgomery, J.C.; Venter, P.J.; Tashian, R.E.; Hewett-Emmett, D.  
 Nucleic Acids Res. 15, 4687, 1987  
 A/Title: Nucleotide sequence of human liver carbonic anhydrase II cDNA.  
 A/Reference number: I37214; MUID:87231043; PMID:3108857  
 A/Accession: I37214  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-260 <RE3>  
 A/Cross-references: EMBL:Y00339; NID:G29586; PIDDN:CAA68426.1; PID:G29587  
 A/Experimental source: liver  
 A/Note: submitted to the EMBL/GenBank/DBJ databases by David Hewett-Emmett 01-JUL-19  
 Am. J. Hum. Genet. 49, 1082-1090, 1991  
 A/Title: Carbonic anhydrase II deficiency syndrome in a Belgian family is caused by a  
 e.

A/Reference number: I51863; MUID:92026087; PMID:1928091  
 A/Accession: I51863  
 A/Status: translation not shown; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-260 <RES>  
 A/Cross-references: GB:M77181; NID:G197778; PIDDN:AAA51909.1; PID:G197780  
 A/Note: the complete nucleotide sequence is not shown  
 R/Hu, P.Y.; Ernst, A.R.; Sly, W.S.; Venter, P.J.; Skaggs, L.A.; Tashian, R.E.  
 Am. J. Hum. Genet. 54, 602-608, 1994  
 A/Title: Carbonic anhydrase II deficiency: single-base deletion in exon 7 is the pred  
 A/Reference number: I51871; MUID:94157074; PMID:8128957  
 A/Accession: I51871  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 225-242 <RE2>  
 A/Cross-references: GB:G59526; NID:G545850; PIDDN:AAA30170.1; PID:G545851  
 A/Note: wild type shown; mutant contains frameshift after residue 226  
 R/Eriksson, A.B.; Jones, T.A.; Liljas, A.  
 submitted to the Brookhaven Protein Data Bank, February 1989  
 A/Reference number: A50085; PDB:1CA2  
 A/Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 4-259  
 R/Liljas, A.; Kannan, K.K.; Bergsten, P.C.; Waara, I.; Fridborg, K.; Strandberg, B.;  
 Nature New Biol. 235, 131-137, 1972  
 A/Title: Crystal structure of human carbonic anhydrase C.  
 A/Reference number: A93404; MUID:72111787; PMID:4621826  
 A/Contents: annotation; X-ray crystallography, 2.0 angstroms  
 A/Note: other residues at the active site are His-64, Asn-67, Tyr-127, Leu-197, Thr-1;  
 C/Genetics:  
 A/Gene: GDB:CA2  
 A/Cross-references: GDB:I19739; OMIM:259730  
 A/Map position: Eq13-8q22.1  
 A/Introns: 12/1; 78/1; 117/3; 146/3; 169/3; 221/3  
 C/Function:

A:Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide  
 A:Note: this form is expressed in erythrocytes and other tissues; deficiency of this for  
 C:Superfamily: carbonate dehydratase, carbonic anhydrase homology  
 C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; metalloprotein; mono  
 F/2-260/Product: carbonate dehydratase II #status experimental <MAT>  
 F/5-250/Domain: carbonic anhydrase homology <CAN>  
 F/2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental  
 F/94,96,119/Binding site: zinc (His) #status experimental

Query Match 15.7%; Score 379.5; DB 1; Length 260;  
 Best Local Similarity 34.5%; Pred. No. 3.3e-17;  
 Matches 90; Conservative 42; Mismatches 114; Indels 15; Gaps 7;

QY 140 HMRYG---GDPMPRVSPACAGRFQSPVDIRPOLAFCALRPPELLGQPLPPLRLR 196  
 DB 4 HMGYGKNGPBEHMHKDFIAKGERQSPVIDHTKYPDLKPSLV---SYDQATSLRL 60  
 QY 197 NNGHSVQLTPPGLEMAL---GP-GREYRALQLHMGAGRPGESEHTVEGRFPALIV 252  
 DB 61 NNGHAFNVEFDDSCDAVKKGSPDLGTYRLIQFHFHMGSLDQGSSEHTVKKKVADELHL 120  
 QY 253 VHLSTAFARVDEALGRPGGLAVLAFLBEGPEENSAVEQLSLREELAESESTQVPGLD 312  
 DB 121 VHMNTKYGDGKAVAGQPGDLAVLGIFLKVG--SAKRGQKQVVDLSIKTKGKADFTNFD 199  
 QY 313 ISALLPDSFRFYQEGSLITTPCAQGVITVFNQTMLSAKQL---HTLSDTLWPGDS 369  
 DB 180 PGLLPESLD-YMTYPSGLTTPPLECTVITVLEKPSISSEQVLEKPKRLNFGEGEPFR 238  
 QY 370 RLQINFRAOTPLNGRVIEASFP 390  
 DB 239 LNVDMWRPAPQPLKRGVYKASF 259

## RESULT 12

CRSB2  
 Carbonate dehydratase (EC 4.2.1.1) II - sheep (tentative sequence)  
 N:Alternate names: carbonic anhydrase II  
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
 C>Date: 31-Mar-1981 #sequence\_revision 31-Mar-1981 #text\_change 31-Mar-2000  
 C/Accession: A01145  
 R/Tanis, R.U.; Ferrell, R.E.; Tashian, R.E.  
 Biochim. Biophys. Acta 371, 534-548, 1974  
 A>Title: Amino acid sequence of sheep carbonic anhydrase C.  
 A:Reference number: A90598; MUID:75054988; PMID:4215456  
 A/Accession: A01145

A:Molecule type: protein  
 A:Residues: 1-259 <CAN>  
 R/Mallet, B.; Gallan, J.M.; Sciaky, M.; Laurent, G.; Charrel, M.  
 Biochim. Biophys. Acta 576, 290-304, 1979  
 A>Title: Formes moléculaires multiples de l'anhydrase carbonique erythrocytaire ovine.  
 A:Reference number: A90628; MUID:79145542; PMID:106895  
 A/Contents: annotation  
 A:Note: one minor and three major forms were isolated chromatographically. One of these  
 C:Superfamily: carbonate dehydratase, carbonic anhydrase homology  
 C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc  
 F/4-258/Domain: carbonic anhydrase homology <CAN>  
 F/1/Modified site: acetylated amino end (Ser) #status experimental  
 F/93,95,118/Binding site: zinc (His) #status predicted

Query Match 15.6%; Score 379; DB 1; Length 259;  
 Best Local Similarity 35.5%; Pred. No. 3.5e-17;  
 Matches 93; Conservative 41; Mismatches 112; Indels 16; Gaps 8;

QY 140 HMRYG---GDPMPRVSPACAGRFQSPVDIRPOLAFCALRPPELLGQPLPPLRLR 196  
 DB 3 HMGYGKNGPBEHMHKDFIADGERSQSPVIDTKAVDPDLKPLALLYEQ---AASRRW 59  
 QY 197 NNGHSVQLTPPGLEMAL---GP-GREYRALQLHMGAGRPGESEHTVEGRFPALIV 252  
 DB 60 NNGHSFNVVEFDDSCDAVKKGSPDLGTYRLIQFHFHMGSLDQGSSEHTVKKKVADELHL 119  
 QY 253 VHLSTAFARVDEALGRPGGLAVLAFLBEGPEENSAVEQLSLREELAESESTQVPGLD 312

DB 120 VHMNTKYGDGKAVAGQPGDLAVLGIFLKVG--DANPALQKLVLDLSIKTKGSAADPNFD 178  
 QY 313 ISALLPDSFRFYQEGSLITTPCAQGVITVFNQTMLSAKQL---HTLSDTLWPGDS 369  
 DB 179 PSSILKRLALN-YMTYPSGLTTPPLECTVITVLEKPSISSEQVLEKPKRLNFGEGEPFR 237  
 QY 370 RLQINFRAOTPLNGRVIEASFP 391  
 DB 238 LNVDMWRPAPQPLKRGVYKASF 258

## RESULT 13

CRSB2  
 Carbonate dehydratase (EC 4.2.1.1) II - rabbit (tentative sequence)  
 N:Alternate names: carbonic anhydrase II  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C>Date: 31-Mar-1981 #sequence\_revision 31-Mar-1981 #text\_change 31-Mar-2000  
 C/Accession: A01142  
 R/Ferrell, R.E.; Stroup, S.K.; Tanis, R.J.; Tashian, R.E.  
 Biochim. Biophys. Acta 533, 1-11, 1978  
 A>Title: Amino acid sequence of rabbit carbonic anhydrase II.  
 A:Reference number: A01142; MUID:78144871; PMID:416851  
 A/Accession: A01142  
 A:Molecule type: protein  
 A:Residues: 1-259 <FER>  
 A:Note: 203-Glu was also found  
 C:Superfamily: carbonate dehydratase, carbonic anhydrase homology  
 C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc  
 F/4-258/Domain: carbonic anhydrase homology <CAN>  
 F/1/Modified site: acetylated amino end (Ser) #status experimental  
 F/93,95,118/Binding site: zinc (His) #status predicted

Query Match 15.4%; Score 374.5; DB 1; Length 259;  
 Best Local Similarity 34.5%; Pred. No. 6.8e-17;  
 Matches 90; Conservative 42; Mismatches 114; Indels 15; Gaps 8;

QY 140 HMRYG---GDPMPRVSPACAGRFQSPVDIRPOLAFCALRPPELLGQPLPPLRLR 196  
 DB 3 HMGYGKNGPBEHMHKDFIADGERSQSPVIDTKAVDPDLKPLALLYEQ---AASRR 59  
 QY 197 NNGHSVQLTPPGLEMAL---GP-GREYRALQLHMGAGRPGESEHTVEGRFPALIV 252  
 DB 60 NNGHSFNVVEFDDSHDKSLKEGPLEGTYRLIQFHFHMGSLDQGSSEHTVKKKVADELHL 119  
 QY 253 VHLSTAFARVDEALGRPGGLAVLAFLBEGPEENSAVEQLSLREELAESESTQVPGLD 312  
 DB 120 VHMNTKYGDGKAVAGQPGDLAVLGIFLKVG--SATPGLQKVVDTLSIKTKGKSVDFNFD 178  
 QY 313 ISALLPDSFRFYQEGSLITTPCAQGVITVFNQTMLSAKQL---HTLSDTLWPGDS 369  
 DB 179 PGLLPESLD-YMTYPSGLTTPPLECTVITVLEKPSISSEQVLEKPKRLNFGEGEPFR 237  
 QY 370 RLQINFRAOTPLNGRVIEASFP 390  
 DB 238 PNVDMWRPAPQPLKRGVYKASF 258

## RESULT 14

JH0527  
 Carbonate dehydratase (EC 4.2.1.1) II - rat  
 N:Alternate names: carbonic anhydrase II  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 22-Jun-1999  
 C/Accession: JH0527; S20114  
 R/Stoll, C.A.; McGowan, M.H.; Heim, R.A.; Varia, M.; Neubauer, J.A.  
 Gene 109, 265-267, 1991  
 A>Title: Nucleotide sequence of a cDNA encoding rat brain carbonic anhydrase II and  
 A:Reference number: JH0527; MUID:92112053; PMID:1765271  
 A/Accession: JH0527  
 A:Molecule type: mRNA  
 A:Residues: 1-260 <STO>  
 A/Cross-references: EMBL:X58294; NID:955837; PID:CAA41227.1; PID:955838

A:Experimental source: brain  
 C:Comment: Carbonate dehydratase is a monomeric zinc metalloenzyme that catalyzes the re  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C:Keywords: carbon-oxygen lyase; hydro-lyase; zinc  
 F:5-259/Domain: carbonic anhydrase homology <CAH>

Query Match 15.2%; Score 366.5; DB 2; Length 260;  
 Best Local Similarity 34.1%; Pred. No. 1,7e-16;  
 Matches 89; Conservative 41; Mismatches 116; Indels 15; Gaps 7;

QY 140 HMY--GSDPWRVSPACGRFQSPVDIRPOLAFCPALRPLELIGFQLPELRLR 196  
 DB 4 HMGYSKNGPENHMKERPIANGDROSPVIDTGAQHDPSLOPLLIC--YDKVASKSIV 60  
 QY 197 NNGHSVQLTPGLEMAL--GP-GREYRALQHLHWGAAGRPGSEHTVEGHRPPAETHV 252  
 DB 61 NNGHSFVVEPDDSDQDFVAKGELSGSYRLIQPHFWGSSDQGSSEHTVKKYNAEDLH 120  
 QY 253 VHSSTAFARVDEALGRPGGLAVLAFLBEPENSAVEQLRLBIEAEGSETQVPGLD 312  
 DB 121 VHWNTKYGDFGKAVQHPDGLAVLGIFLKI GP-ASQGLQKITEALHSIKTKGRAPAFANFD 179  
 QY 313 ISALPDSFRRFYQYEGSLTTPCAQGVITVFNQVMTLSAKQL--HTLSDTLMGPGDS 369  
 DB 180 PCSLLPGLND-VWYTPGSLTTPPLLECVTWIVLKEPTVSSQMSHFRKLNFSSEGAEE 238  
 QY 370 RLQINFRATQPLNGRVLEASF 390  
 DB 239 LMDVNMWRPAQLNKRKIKASF 259

## RESULT 15

CRMOIR  
 carbonate dehydratase (EC 4.2.1.1) I - rhesus macaque (tentative sequence)  
 N:Alternate names: carbonic anhydrase I  
 C:Species: Macaca mulatta (rhesus macaque)  
 C>Date: 07-May-1981 #sequence\_revision 07-May-1981 #text\_change 31-Mar-2000  
 C/Accession: A01139  
 R/Henriksson, D.; Tanis, R.U.; Tashian, R.E.  
 Biochem. Biophys. Res. Commun. 96, 135-142, 1980  
 A/Title: The amino acid sequence of carbonic anhydrase I from the Rhesus macaque.  
 A/Reference number: A01139; NID:81062409; PMID:6776950  
 A/Accession: A01139  
 A:Molecule type: protein  
 A/Residues: 1-260 <RHN>  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc  
 F:5-260/Domain: carbonic anhydrase homology <CAH>  
 F:1/Modified site: acetylated amino end (Ala) #status experimental  
 F:94,96,119/Binding site: zinc (His) #status predicted

Query Match 15.1%; Score 366.5; DB 1; Length 260;

Best Local Similarity 32.6%; Pred. No. 2.2e-16;  
 Matches 87; Conservative 48; Mismatches 109; Indels 23; Gaps 8;

QY 132 DKEGDDQSHWRYGSDPPWPFSPACGRFQSPVDIRPOLAFCPALRPLELIGFQLPLP 191  
 DB 9 DKNGPEQ-----MSKLYPIANGNNQSPVDIKTSEAKHDITLKRISV---SYNPAT 55  
 QY 192 ELRLKNGHGVQLTPGLEMAL--GP-GREYRALQHLHWGAAGRPGSEHTVEGHRFP 247  
 DB 56 AKSIIINGSHFHNFEFNDNRSVLKGPFSDSRLQFHFHWGSSNEYGSEHTVDGVKYS 115  
 QY 248 AEIHVYHLSTA-PARVDEALGRPGGLAVLAFLBEPENSAVEQLRLBIEAEGSET 306  
 DB 116 SELHIYHMNSAKYSLSLAENSKADGLAVIGVLMKVG-EANPKLQVLDALHAIKTKGR 174  
 QY 307 QVPEGLISALLPDSFRRFYQYEGSLTTPCAQGVITVFNQVMTLSAKQL--HTLSDTL 363  
 DB 175 PFTNFPSTLLPGLD-FWYSGSLTHPLPLYSVTWICKESISVSSQLAQFRSLISNV 233  
 QY 364 WGPDSRLQINFRATQPLNGRVLEASF 390

DB 234 EGSNFIQIRNNRFTQPLKGRIVRAS 260

Search completed: March 5, 2004, 09:07:53  
 Job time: 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 09:03:00 ; Search time 22 seconds  
(without alignments)  
1077.105 Million cell updates/sec

Title: US-09-967-237-2

Perfect score: 2424

Sequence: 1 MAPLCSPWLPPLIPAPAPG.....RRGKGVSVRPAEVAETGA 459

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgm2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgm2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
3: /cgm2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
4: /cgm2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
5: /cgm2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
6: /cgm2\_6/prodata/2/1aa/5A\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2424	100.0	459	2	US-08-481-658B-2
2	2424	100.0	459	2	US-08-477-504A-2
3	2424	100.0	459	2	US-08-486-756A-2
4	2424	100.0	459	2	US-08-485-862B-2
5	2424	100.0	459	2	US-08-487-739-2
6	2424	100.0	459	2	US-08-487-077A-2
7	2424	100.0	459	2	US-08-485-863A-2
8	2424	100.0	459	2	US-08-485-049D-2
9	2424	100.0	459	2	US-09-178-115-2
10	2424	100.0	459	2	US-09-177-776-2
11	2227	91.9	422	3	US-08-335-469-2
12	2007	82.8	377	3	US-08-787-739-87
13	2007	82.8	377	3	US-09-178-115-87
14	2007	82.8	377	3	US-09-177-776-87
15	1398.5	57.7	429	1	US-07-664-589-2
16	1398.5	57.7	429	5	PCT-US83-02024-2
17	1370	56.5	257	3	US-08-787-739-51
18	1370	56.5	257	3	US-09-178-115-51
19	1370	56.5	257	3	US-09-177-776-51
20	1364	56.3	256	2	US-08-481-658B-51
21	1364	56.3	256	2	US-08-477-504A-51
22	1364	56.3	256	2	US-08-486-756A-51
23	1364	56.3	256	2	US-08-485-862B-51
24	1364	56.3	256	2	US-08-487-077A-51
25	1364	56.3	256	3	US-08-485-863A-51
26	1364	56.3	256	3	US-08-485-049D-51
27	884	36.5	170	2	US-08-481-658B-54

28	884	36.5	170	2	US-08-477-504A-54	Sequence 54, Appl
29	884	36.5	170	2	US-08-486-756A-54	Sequence 54, Appl
30	884	36.5	170	2	US-08-485-862B-54	Sequence 54, Appl
31	884	36.5	170	3	US-08-487-077A-54	Sequence 54, Appl
32	884	36.5	170	3	US-08-485-863A-54	Sequence 54, Appl
33	884	36.5	170	3	US-08-485-049D-54	Sequence 54, Appl
34	537.5	22.2	354	1	US-08-276-919-2	Sequence 2, Appl
35	537.5	22.2	354	1	US-08-776-088-2	Sequence 2, Appl
36	537.5	22.2	354	1	US-08-776-088-6	Sequence 2, Appl
37	537.5	22.2	354	3	US-09-325-320-2	Sequence 2, Appl
38	537.5	22.2	354	3	US-09-585-109-2	Sequence 2, Appl
39	537.5	22.2	354	5	PCT-US95-09145A-2	Sequence 2, Appl
40	537.5	22.2	354	5	PCT-US95-09145A-6	Sequence 6, Appl
41	535	22.1	325	1	US-08-276-919-4	Sequence 4, Appl
42	535	22.1	325	5	US-08-776-088-4	Sequence 4, Appl
43	535	22.1	325	5	PCT-US95-09145A-4	Sequence 4, Appl
44	533	22.0	325	1	US-08-276-919-13	Sequence 13, Appl
45	533	22.0	325	1	US-08-776-088-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-2  
Sequence 2, Application US/08481658B  
Patent No. 5955075  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30 (ENO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,658B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO.: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 459 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: First 37 amino acids represent  
signal peptide, and remaining amino acids  
DESCRIPTION: represent mature protein  
US-08-481-658B-2  
Query Match 100.0%; Score 2424; DB 2; Length 459;  
Best Local Similarity 100.0%; Pred. No. 8,6e-197;



PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/260,190  
 FILING DATE: 15-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lauder, Leona L.  
 REGISTRATION NUMBER: 30,863  
 REFERENCE/DOCKET NUMBER: D-0021.3C  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-435-2034  
 TELEFAX: 415-435-0727  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 459 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 DESCRIPTION: First 37 amino acids represent  
 signal peptide, and remaining amino acids  
 represent mature protein  
 US-08-486-756A-2

Query Match 100.0%; Score 2424; DB 2; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-197;  
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSMPLPLIPAPAGLTIVQLLSLLLMVHPQLPRMOWDSPLGGSSGDDPL 60  
 DB 1 MAPLCPSMPLPLIPAPAGLTIVQLLSLLLMVHPQLPRMOWDSPLGGSSGDDPL 60  
 QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 QY 121 DPQEPQNNARHDKEDDQSHWRYGDDPPWPRVSPACAGRFQSPVDIRPOLAFCFALRPL 180  
 DB 121 DPQEPQNNARHDKEDDQSHWRYGDDPPWPRVSPACAGRFQSPVDIRPOLAFCFALRPL 180  
 QY 181 ELIGFQLPLPELRIRNNHSHVQLTPPGLEVALGPGREYRALQLHLMWGAAGRPGSEHT 240  
 DB 181 ELIGFQLPLPELRIRNNHSHVQLTPPGLEVALGPGREYRALQLHLMWGAAGRPGSEHT 240  
 QY 241 VEGHRPFAIHVHLSFAFARVDEALGPGGLAVLAFLBEGPEENSAVYEQLSRLEIEA 300  
 DB 241 VEGHRPFAIHVHLSFAFARVDEALGPGGLAVLAFLBEGPEENSAVYEQLSRLEIEA 300  
 QY 301 EEGSETQVPGLDISALLPSDFSRFYQYEGSLTTPCAQGVITWVFNQVWMLSAKQHTLS 360  
 DB 301 EEGSETQVPGLDISALLPSDFSRFYQYEGSLTTPCAQGVITWVFNQVWMLSAKQHTLS 360  
 QY 361 DTLWGGDSRLQINFRATQPLNGRVIEASFPAGVDSPPRAAPVQNSCLAAGDILALVF 420  
 DB 361 DTLWGGDSRLQINFRATQPLNGRVIEASFPAGVDSPPRAAPVQNSCLAAGDILALVF 420  
 QY 421 GLIFAVTSVAFVQMRORHRRGTGVSYPRAEVAETGA 459  
 DB 421 GLIFAVTSVAFVQMRORHRRGTGVSYPRAEVAETGA 459

RESULT 4  
 US-08-485-862B-2  
 Sequence 2, Application US/08485862B  
 Patent No. 5989838  
 GENERAL INFORMATION:  
 APPLICANT: Zavada, Jan  
 APPLICANT: Pastorekova, Silvia  
 APPLICANT: Pastorek, Jaromir  
 TITLE OF INVENTION: MN Gene and Protein  
 NUMBER OF SEQUENCES: 86  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Leona L. Lauder  
 STREET: 6 Mariposa Court  
 CITY: Tiburon

STATE: California  
 COUNTRY: USA  
 ZIP: 94920  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/485,862B  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/477,504  
 FILING DATE: 07-JUN-1995  
 APPLICATION NUMBER: US 08/260,190  
 FILING DATE: 15-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lauder, Leona L.  
 REGISTRATION NUMBER: 30,863  
 REFERENCE/DOCKET NUMBER: D-0021.3D  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-435-2034  
 TELEFAX: 415-435-0727  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 459 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 DESCRIPTION: First 37 amino acids represent  
 signal peptide, and remaining amino acids  
 represent mature protein  
 US-08-485-862B-2

Query Match 100.0%; Score 2424; DB 2; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-197;  
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSMPLPLIPAPAGLTIVQLLSLLLMVHPQLPRMOWDSPLGGSSGDDPL 60  
 DB 1 MAPLCPSMPLPLIPAPAGLTIVQLLSLLLMVHPQLPRMOWDSPLGGSSGDDPL 60  
 QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 QY 121 DPQEPQNNARHDKEDDQSHWRYGDDPPWPRVSPACAGRFQSPVDIRPOLAFCFALRPL 180  
 DB 121 DPQEPQNNARHDKEDDQSHWRYGDDPPWPRVSPACAGRFQSPVDIRPOLAFCFALRPL 180  
 QY 181 ELIGFQLPLPELRIRNNHSHVQLTPPGLEVALGPGREYRALQLHLMWGAAGRPGSEHT 240  
 DB 181 ELIGFQLPLPELRIRNNHSHVQLTPPGLEVALGPGREYRALQLHLMWGAAGRPGSEHT 240  
 QY 241 VEGHRPFAIHVHLSFAFARVDEALGPGGLAVLAFLBEGPEENSAVYEQLSRLEIEA 300  
 DB 241 VEGHRPFAIHVHLSFAFARVDEALGPGGLAVLAFLBEGPEENSAVYEQLSRLEIEA 300  
 QY 301 EEGSETQVPGLDISALLPSDFSRFYQYEGSLTTPCAQGVITWVFNQVWMLSAKQHTLS 360  
 DB 301 EEGSETQVPGLDISALLPSDFSRFYQYEGSLTTPCAQGVITWVFNQVWMLSAKQHTLS 360  
 QY 361 DTLWGGDSRLQINFRATQPLNGRVIEASFPAGVDSPPRAAPVQNSCLAAGDILALVF 420  
 DB 361 DTLWGGDSRLQINFRATQPLNGRVIEASFPAGVDSPPRAAPVQNSCLAAGDILALVF 420  
 QY 421 GLIFAVTSVAFVQMRORHRRGTGVSYPRAEVAETGA 459  
 DB 421 GLIFAVTSVAFVQMRORHRRGTGVSYPRAEVAETGA 459

## RESULT 5

US-08-787-739-2  
Sequence 2, Application US/08787739  
Patent No. 6027887  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jarmir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 96  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street, Suite 610  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 459 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
DESCRIPTION: First 37 amino acids represent  
signal peptide, and remaining amino acids  
DESCRIPTION: represent mature protein  
US-08-787-739-2

Query Match 100.0%; Score 2424; DB 3; Length 459;  
Best Local Similarity 100.0%; Pred. No. 8.6e-197;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSFWPLFLIPAPAGITVQLLSLLLMFVHPORLPFMOEDSPVGGSSGSDPL 60  
DB 1 MAPLCPSFWPLFLIPAPAGITVQLLSLLLMFVHPORLPFMOEDSPVGGSSGSDPL 60  
QY 61 GEEDLPSEEDSPREEDPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120

DB 61 GEEDLPSEEDSPREEDPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
QY 121 DPOEPQNNARXDEGDDQSHWRYGDPMPWRVSPACAGRFQSPVDIRPOLAFCPALRPL 180  
DB 121 DPOEPQNNARXDEGDDQSHWRYGDPMPWRVSPACAGRFQSPVDIRPOLAFCPALRPL 180  
QY 181 ELIGFQPLPPELRLNNHSHVQLTPCLDENALPGREYRLQHLHMGAAAGRGSEHT 240  
DB 181 ELIGFQPLPPELRLNNHSHVQLTPCLDENALPGREYRLQHLHMGAAAGRGSEHT 240  
QY 241 VEGHPPAEIHVHLSSTAPRVDEALGRGGLAVLAFLBEGPEENSAYEQLSRLEBIA 300  
DB 241 VEGHPPAEIHVHLSSTAPRVDEALGRGGLAVLAFLBEGPEENSAYEQLSRLEBIA 300  
QY 301 BEGSETQVPGDLISALLPDSFRTYFEGSLTTPCAQGVITWVNTVMISAKQLHTLS 360  
DB 301 BEGSETQVPGDLISALLPDSFRTYFEGSLTTPCAQGVITWVNTVMISAKQLHTLS 360  
QY 361 DTLMGDSRLQINFRATQPLNGRVTEASFPAQVDSPPRAEPVOLNSCLAAGDILALVF 420  
DB 361 DTLMGDSRLQINFRATQPLNGRVTEASFPAQVDSPPRAEPVOLNSCLAAGDILALVF 420  
QY 421 GLIFAVTSVAFVQWRQRHRTGKGVSYRPAEVAETGA 459  
DB 421 GLIFAVTSVAFVQWRQRHRTGKGVSYRPAEVAETGA 459

## RESULT 6

US-08-487-077A-2  
Sequence 2, Application US/08487077A  
Patent No. 6069242

GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jarmir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,077A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3H  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 459 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: First 37 amino acids represent  
signal peptide, and remaining amino acids



DESCRIPTION: represent mature protein  
US-08-487-077A-2

Query Match 100.0%; Score 2424; DB 3; Length 459;  
Best Local Similarity 100.0%; Pred. No. 8,66-197;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSWMLPILIPAPAGLTQVLLSLILMVPVHQRLLPRMDESPLOGSSGSDPL 60  
DB 1 MAPLCPSWMLPILIPAPAGLTQVLLSLILMVPVHQRLLPRMDESPLOGSSGSDPL 60  
QY 61 GEEDLPSEEDSPREEDPPGEBDLPGEEDLPGEEDLPVKKPSEEGSLKLEDLPTVEAPG 120  
DB 61 GEEDLPSEEDSPREEDPPGEBDLPGEEDLPGEEDLPVKKPSEEGSLKLEDLPTVEAPG 120  
QY 121 DPOEPONNAHRDKEGDDQSHWRVYGGDPMPWRVSPACAGRFOSPVDIRPOLAFCPALRPL 180  
DB 121 DPOEPONNAHRDKEGDDQSHWRVYGGDPMPWRVSPACAGRFOSPVDIRPOLAFCPALRPL 180  
QY 181 ELIGFQLPPLPELRRLNNGHSVQLTPPGLEMALGPGEYRALQHLHWGAAGRPGSEHT 240  
DB 181 ELIGFQLPPLPELRRLNNGHSVQLTPPGLEMALGPGEYRALQHLHWGAAGRPGSEHT 240  
QY 241 VEGHRFPPEIHVHVLSTAFARVDEALGRGGLAVLAFLBEGPEENSAVEQLLSRLEIA 300  
DB 241 VEGHRFPPEIHVHVLSTAFARVDEALGRGGLAVLAFLBEGPEENSAVEQLLSRLEIA 300  
QY 301 EBGSETOVPGDLISALLPSDFSRFYQYEGSLTPPCAQGVITWTFNQVWLSAQVHTLS 360  
DB 301 EBGSETOVPGDLISALLPSDFSRFYQYEGSLTPPCAQGVITWTFNQVWLSAQVHTLS 360  
QY 361 DTLMGPGDSRLQINFRAQPLNGRVIEASFPAGVSSPRAEPVQVNSCLAAGDILALVF 420  
DB 361 DTLMGPGDSRLQINFRAQPLNGRVIEASFPAGVSSPRAEPVQVNSCLAAGDILALVF 420  
QY 421 GLIFAVTSVAFLVQMRQRHRRGTGKGVSYRAEVAETGA 459  
DB 421 GLIFAVTSVAFLVQMRQRHRRGTGKGVSYRAEVAETGA 459

## RESULT 7

US-08-485-863A-2  
Sequence 2, Application US/08485863A  
Patent No. 6093548  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,863A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3G

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 459 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: First 37 amino acids represent  
signal peptide, and remaining amino acids  
DESCRIPTION: represent mature protein  
US-08-485-863A-2

Query Match 100.0%; Score 2424; DB 3; Length 459;  
Best Local Similarity 100.0%; Pred. No. 8,66-197;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSWMLPILIPAPAGLTQVLLSLILMVPVHQRLLPRMDESPLOGSSGSDPL 60  
DB 1 MAPLCPSWMLPILIPAPAGLTQVLLSLILMVPVHQRLLPRMDESPLOGSSGSDPL 60  
QY 61 GEEDLPSEEDSPREEDPPGEBDLPGEEDLPGEEDLPVKKPSEEGSLKLEDLPTVEAPG 120  
DB 61 GEEDLPSEEDSPREEDPPGEBDLPGEEDLPGEEDLPVKKPSEEGSLKLEDLPTVEAPG 120  
QY 121 DPOEPONNAHRDKEGDDQSHWRVYGGDPMPWRVSPACAGRFOSPVDIRPOLAFCPALRPL 180  
DB 121 DPOEPONNAHRDKEGDDQSHWRVYGGDPMPWRVSPACAGRFOSPVDIRPOLAFCPALRPL 180  
QY 181 ELIGFQLPPLPELRRLNNGHSVQLTPPGLEMALGPGEYRALQHLHWGAAGRPGSEHT 240  
DB 181 ELIGFQLPPLPELRRLNNGHSVQLTPPGLEMALGPGEYRALQHLHWGAAGRPGSEHT 240  
QY 241 VEGHRFPPEIHVHVLSTAFARVDEALGRGGLAVLAFLBEGPEENSAVEQLLSRLEIA 300  
DB 241 VEGHRFPPEIHVHVLSTAFARVDEALGRGGLAVLAFLBEGPEENSAVEQLLSRLEIA 300  
QY 301 EBGSETOVPGDLISALLPSDFSRFYQYEGSLTPPCAQGVITWTFNQVWLSAQVHTLS 360  
DB 301 EBGSETOVPGDLISALLPSDFSRFYQYEGSLTPPCAQGVITWTFNQVWLSAQVHTLS 360  
QY 361 DTLMGPGDSRLQINFRAQPLNGRVIEASFPAGVSSPRAEPVQVNSCLAAGDILALVF 420  
DB 361 DTLMGPGDSRLQINFRAQPLNGRVIEASFPAGVSSPRAEPVQVNSCLAAGDILALVF 420  
QY 421 GLIFAVTSVAFLVQMRQRHRRGTGKGVSYRAEVAETGA 459  
DB 421 GLIFAVTSVAFLVQMRQRHRRGTGKGVSYRAEVAETGA 459

## RESULT 8

US-08-485-049D-2  
Sequence 2, Application US/08485049D  
Patent No. 6204370  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```

1      SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
2      CURRENT APPLICATION DATA:
3      APPLICATION NUMBER: US/08/465,049D
4      FILING DATE: 07-JUN-1995
5      CLASSIFICATION: 435
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER: US 08/260,190
8      FILING DATE: 15-JUN-1994
9      ATTORNEY/AGENT INFORMATION:
10     NAME: Lauder, Leona L.
11     REGISTRATION NUMBER: 30,863
12     REFERENCE/DOCKET NUMBER: D-0021.3E
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE: 415-981-2034
15     TELEFAX: 415-981-0332
16     INFORMATION FOR SEQ ID NO: 2:
17     SEQUENCE CHARACTERISTICS:
18     LENGTH: 459 amino acids
19     TYPE: amino acid
20     STRANDEDNESS:
21     TOPOLOGY: linear
22     MOLECULE TYPE: protein
23     DESCRIPTION: First 37 amino acids represent
24     DESCRIPTION: signal peptide, and remaining amino acids
25     DESCRIPTION: represent mature protein
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
```

## RESULT 10

US-09-177-776-2

Sequence 2, Application US/09177776A

Patent No. 6237051

GENERAL INFORMATION:

APPLICANT: Zavada, Jan

APPLICANT: Pastorekova, Silvia

APPLICANT: Pastorek, Jaromir

TITLE OF INVENTION: MN Gene and Protein

FILE REFERENCE: D-0021 SA

CURRENT APPLICATION NUMBER: US/09/177,776A

EARLIER FILING DATE: 1998-10-23

EARLIER APPLICATION NUMBER: 08/787,739

EARLIER FILING DATE: 1997-01-24

EARLIER APPLICATION NUMBER: 08/485,049

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/486,756

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/477,504

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/481,658

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/485,862

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/485,863

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/487,077

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/260,190

EARLIER FILING DATE: 1994-06-15

EARLIER APPLICATION NUMBER: 08/177,093

EARLIER FILING DATE: 1993-12-30

EARLIER APPLICATION NUMBER: 07/964,589

EARLIER FILING DATE: 1992-10-21

EARLIER APPLICATION NUMBER: PV-709-92

EARLIER FILING DATE: 1992-03-11

NUMBER OF SEQ ID NOS: 116

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 459

TYPE: PRT

ORGANISM: HUMAN

US-09-177-776-2

Query Match 100.0%; Score 2424; DB 3; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-197; Indels 0; Gaps 0;  
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NAFPCSPWLPPLIPAPAPAGLTIVQLLSLILMFPVHQRPRMOWEDSPILGGSSGDDPL 60  
 DB 1 MAFPCSPWLPPLIPAPAPAGLTIVQLLSLILMFPVHQRPRMOWEDSPILGGSSGDDPL 60  
 QY 61 GSEEDLPSEEDSREDDPGGEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 DB 61 GSEEDLPSEEDSREDDPGGEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 QY 121 DQEPQONNAHDKEDDQSHWRYGDDPFWPRVSPACAGRFQSPVDIRPOLAFCPALRPL 180  
 DB 121 DQEPQONNAHDKEDDQSHWRYGDDPFWPRVSPACAGRFQSPVDIRPOLAFCPALRPL 180  
 QY 181 ELIGFOLPPLPELRIRNNGSHVQTLTPGLGEMALGPGREYRALQHLHWAAGRPGSEHT 240  
 DB 181 ELIGFOLPPLPELRIRNNGSHVQTLTPGLGEMALGPGREYRALQHLHWAAGRPGSEHT 240  
 QY 241 VEGHRFPAEIHVHLSTAFARVDEALGRPGGLAVLAFAFLEEGPENSAVYQLSLRLEEA 300  
 DB 241 VEGHRFPAEIHVHLSTAFARVDEALGRPGGLAVLAFAFLEEGPENSAVYQLSLRLEEA 300  
 QY 301 EGGSETQVPGDLISALLSSDPGRFYQVHGSILTPPCQGVYVWVFNCTVMSAKQHTLS 360  
 DB 301 EGGSETQVPGDLISALLSSDPGRFYQVHGSILTPPCQGVYVWVFNCTVMSAKQHTLS 360

QY 361 DTLMGPGDSRLQINFRATOPUNGRIEASFPAGVDSPPRAEPVQINSCLAADILATVE 420  
 DB 361 DTLMGPGDSRLQINFRATOPUNGRIEASFPAGVDSPPRAEPVQINSCLAADILATVE 420  
 QY 421 GILFAVTSVAFVQMRQRRGRTGVSYPAPAEVATGA 459  
 DB 421 GILFAVTSVAFVQMRQRRGRTGVSYPAPAEVATGA 459

## RESULT 11

US-08-335-469-2

Sequence 2, Application US/08335469A

Patent No. 6004535

GENERAL INFORMATION:

APPLICANT: Zavada, Jan

APPLICANT: Pastorekova, Silvia

APPLICANT: Pastorek, Jaromir

TITLE OF INVENTION: Methods to Detect and Quantify MN Protein/Polypeptide

FILE REFERENCE: D-0021A

CURRENT APPLICATION NUMBER: US/08/335,469A

EARLIER FILING DATE: 1994-11-07

EARLIER APPLICATION NUMBER: 07/964,589

EARLIER FILING DATE: 1992-10-21

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 422

TYPE: PRT

ORGANISM: Human

US-08-335-469-2

Query Match 91.9%; Score 2227; DB 3; Length 422;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-180; Indels 0; Gaps 0;  
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 QRLPRMOWEDSPILGGSSGDDPLGEEDLPSEEDSPREDDPGGEDLPGEEDLPGEEDLP 97  
 DB 1 QRLPRMOWEDSPILGGSSGDDPLGEEDLPSEEDSPREDDPGGEDLPGEEDLPGEEDLP 97  
 QY 98 VKPSEEGSLKLEEDLPTEAPDPOEPONNAHDKEDDQSHWRYGDDPFWPRVSPACA 157  
 DB 98 VKPSEEGSLKLEEDLPTEAPDPOEPONNAHDKEDDQSHWRYGDDPFWPRVSPACA 157  
 QY 61 VKPSEEGSLKLEEDLPTEAPDPOEPONNAHDKEDDQSHWRYGDDPFWPRVSPACA 120  
 DB 61 VKPSEEGSLKLEEDLPTEAPDPOEPONNAHDKEDDQSHWRYGDDPFWPRVSPACA 120  
 QY 158 GRFQSPVDIRPOLAFCPALRPLLELGFOLPPELRIRNNGSHVQTLTPGLGEMALGPG 217  
 DB 121 GRFQSPVDIRPOLAFCPALRPLLELGFOLPPELRIRNNGSHVQTLTPGLGEMALGPG 180  
 QY 218 REYRALQHLHWAAGRPGSEHTVEGHRFPAEIHVHLSTAFARVDEALGRPGGLAVLA 277  
 DB 181 REYRALQHLHWAAGRPGSEHTVEGHRFPAEIHVHLSTAFARVDEALGRPGGLAVLA 240  
 QY 278 FLEEGPENSAVYQLSLRLEEAEGSETQVPGDLISALLSPDPGRFYQVHGSILTPPCA 337  
 DB 241 FLEEGPENSAVYQLSLRLEEAEGSETQVPGDLISALLSPDPGRFYQVHGSILTPPCA 300  
 QY 338 QGVITVFNCTVMSAKQHTLSDTLMGPGDSRLQINFRATOPUNGRIEASFPAGVDS 397  
 DB 301 QGVITVFNCTVMSAKQHTLSDTLMGPGDSRLQINFRATOPUNGRIEASFPAGVDS 360  
 QY 398 PRAEPVQINSCLAAGITLAVGGLPAYSVAFLVQMRQRRGRTGVSYPAPAEVAT 457  
 DB 361 PRAEPVQINSCLAAGITLAVGGLPAYSVAFLVQMRQRRGRTGVSYPAPAEVAT 420  
 QY 458 GA 459  
 DB 421 GA 422

## RESULT 12

US-08-787-739-87

Sequence 87, Application US/08787739

Patent No. 6027887

GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 96  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street, Suite 610  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 87:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 377 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: MN protein extracellular domain  
US-08-787-739-87  
Query Match 82.8%; Score 2007; DB 3; Length 377;  
Best Local Similarity 100.0%; Pred. No. 1.2e-161;  
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

218 REFRAQLHFWGAAGRPSSEHTVEGHRPAETIHYVHSTAPARVDALGRPGGLAVIAA 277  
DB 181 REFRAQLHFWGAAGRPSSEHTVEGHRPAETIHYVHSTAPARVDALGRPGGLAVIAA 240  
QY 278 FLEEGPENSAVEQLSLRLEELIAEGSETOVGLDISALPDEFERYQYEGSLTTPPCA 337  
DB 241 FLEEGPENSAVEQLSLRLEELIAEGSETOVGLDISALPDEFERYQYEGSLTTPPCA 300  
QY 338 QGVITWYFNQTVWLSAKQHTLSDTLMGSGDSRLQNFRAIPLNGRVIEMSPFAGVSS 397  
DB 301 QGVITWYFNQTVWLSAKQHTLSDTLMGSGDSRLQNFRAIPLNGRVIEMSPFAGVSS 360  
QY 398 PRAAPYQUNSCIAAGD 414  
DB 361 PRAAPYQUNSCIAAGD 377

RESULT 13  
US-09-178-115-87  
Sequence 87, Application US/09178115  
Patent No. 6297041  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
FILE REFERENCE: D-0021.5A  
CURRENT APPLICATION NUMBER: US/09/178,115  
CURRENT FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 09/177,776  
EARLIER FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 08/787,739  
EARLIER FILING DATE: 1997-01-24  
EARLIER APPLICATION NUMBER: 08/485,049  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/486,756  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/477,504  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/481,658  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,862  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,863  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/487,077  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/260,190  
EARLIER FILING DATE: 1994-06-15  
EARLIER APPLICATION NUMBER: 08/177,093  
EARLIER FILING DATE: 1993-12-30  
EARLIER APPLICATION NUMBER: 07/964,589  
EARLIER FILING DATE: 1992-10-21  
EARLIER APPLICATION NUMBER: PV-709-92  
EARLIER FILING DATE: 1992-03-11  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 87  
LENGTH: 377  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-178-115-87  
Query Match 82.8%; Score 2007; DB 3; Length 377;  
Best Local Similarity 100.0%; Pred. No. 1.2e-161;  
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 61 VKKSEEGSLKLEDPTEVARGDPQEPQNNAHKXGDDQSHRYGSDPMPVSPACA 120  
Qy 158 GRQSPVDIRPOLAACPALRPELLELGPQLPELRLRNNGHSVQLTLPGLMALGPG 217  
Db 121 GRQSPVDIRPOLAACPALRPELLELGPQLPELRLRNNGHSVQLTLPGLMALGPG 180  
Qy 218 REYRALQHLHWGAAGRPGSEHTVEGHRFPALIHVHLSTAFARVDALGRPGGLAVLAA 277  
Db 181 REYRALQHLHWGAAGRPGSEHTVEGHRFPALIHVHLSTAFARVDALGRPGGLAVLAA 240  
Qy 278 FLEEGPENSAYEQLSLRLEIAEGSETOVPGDLISALPSPFSRYFYQYEGSLTTPPCA 337  
Db 241 FLEEGPENSAYEQLSLRLEIAEGSETOVPGDLISALPSPFSRYFYQYEGSLTTPPCA 300  
Qy 338 QGVITVFNQTVWLSAKQHLTSLDITLWPGDSRLQINFRATQPLNGRVIETASPPAGVDS 397  
Db 301 QGVITVFNQTVWLSAKQHLTSLDITLWPGDSRLQINFRATQPLNGRVIETASPPAGVDS 360  
Qy 398 PRAEPVQVNSCLAAGD 414  
Db 361 PRAEPVQVNSCLAAGD 377

## RESULT 14

US-09-177-776-87  
Sequence 87, Application US/09177776A  
Patent No. 6297051  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
FILE REFERENCE: D-0021.5A  
CURRENT APPLICATION NUMBER: US/09/177,776A  
CURRENT FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 08/787,739  
EARLIER FILING DATE: 1997-01-24  
EARLIER APPLICATION NUMBER: 08/485,049  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/486,756  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/477,504  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/481,558  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,862  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,863  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/487,077  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/260,190  
EARLIER FILING DATE: 1994-06-15  
EARLIER APPLICATION NUMBER: 08/177,093  
EARLIER FILING DATE: 1993-12-30  
EARLIER APPLICATION NUMBER: 07/964,589  
EARLIER FILING DATE: 1992-10-21  
EARLIER APPLICATION NUMBER: PV-709-92  
EARLIER FILING DATE: 1992-03-11  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 87  
LENGTH: 377  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-177-776-87

Query Match 82.8%; Score 2007; DB 3; Length 377;  
Best Local Similarity 100.0%; Pred. No. 1,2e-161;  
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 38 QRLPRMQDSPLGGSSGSDDDPLGEEDLPSEEDSPREEDPGEEDLPGEEDLPGEEDLPGE 97

Db 1 QRLPRMQDSPLGGSSGSDDDPLGEEDLPSEEDSPREEDPGEEDLPGEEDLPGEEDLPGE 60  
Qy 98 VPKSEEGSLKLEDPTEVARGDPQEPQNNAHKXGDDQSHRYGSDPMPVSPACA 157  
Db 61 VPKSEEGSLKLEDPTEVARGDPQEPQNNAHKXGDDQSHRYGSDPMPVSPACA 120  
Qy 158 GRQSPVDIRPOLAACPALRPELLELGPQLPELRLRNNGHSVQLTLPGLMALGPG 217  
Db 121 GRQSPVDIRPOLAACPALRPELLELGPQLPELRLRNNGHSVQLTLPGLMALGPG 180  
Qy 218 REYRALQHLHWGAAGRPGSEHTVEGHRFPALIHVHLSTAFARVDALGRPGGLAVLAA 277  
Db 181 REYRALQHLHWGAAGRPGSEHTVEGHRFPALIHVHLSTAFARVDALGRPGGLAVLAA 240  
Qy 278 FLEEGPENSAYEQLSLRLEIAEGSETOVPGDLISALPSPFSRYFYQYEGSLTTPPCA 337  
Db 241 FLEEGPENSAYEQLSLRLEIAEGSETOVPGDLISALPSPFSRYFYQYEGSLTTPPCA 300  
Qy 338 QGVITVFNQTVWLSAKQHLTSLDITLWPGDSRLQINFRATQPLNGRVIETASPPAGVDS 397  
Db 301 QGVITVFNQTVWLSAKQHLTSLDITLWPGDSRLQINFRATQPLNGRVIETASPPAGVDS 360  
Qy 398 PRAEPVQVNSCLAAGD 414  
Db 361 PRAEPVQVNSCLAAGD 377

## RESULT 15

US-07-964-589-2  
Sequence 2, Application US/07964589  
Patent No. 5387676  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: Steuart Street Tower, 18th Fl., One Market  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/964,589  
FILING DATE: 19921021  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-777-9257  
TELEFAX: 415-543-4219  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 429 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-964-589-2

Query Match 57.7%; Score 1398.5; DB 1; Length 429;  
Best Local Similarity 69.1%; Pred. No. 4e-110;  
Matches 300; Conservative 12; Mismatches 83; Indels 39; Gaps 7;

```

QY 38 QRLPRMOEDSPICGSSGSGEDDPLGEBDLPSREDSPREDDPGEDDLGCEEDLPGEDL--P 96
Db 1 QRLPRMOEDSPLEELLLGKTMWARRICPVKRIRIHP-ERRITLERIRITL 59
QY 97 EYKPSSEEGSLKLEDLPTVEAPGDPONNARDEKGDQSHRYGSDPWP----- 150
Db 60 KLMPSSEEGSLKLEDLPTVEAPGDPONNARDEKGDQSHRYGSDPWPCCPQPA 119
QY 151 --RVSPPACAGRFQSPVDIRPOLAFC-----PALRPLELGFQJPLPELRNNGH 200
Db 120 RAASSPRWISAPSSPPSAP-----CAPMNSMASSRRSONCACQWP----- 162
QY 201 SVQTLTPGLEMALPGRE--YRALQHLHWGAAGRPQSEHTVEGHRFPALIHVHLSTA 258
Db 163 --QCATDPASWABDGSGRAGVPALQHLHWGAAGRPQSEHTVEGHRFPALIHVHLSTA 220
QY 259 FARVDEALGRPGGLAVLAFLBEGPEENSAYEQLLSRLEIAEBSFTQVPGDLSALLP 318
Db 221 FARVDEALGRPGGLAVLAFWRARRAKKTIVSEQLLSRLEIAEBSFTQVPGDLSALLP 280
QY 319 SPSRXYFOYEGSLTPCAQGVITVFNQTVMLSAKQHLTLSDTLWPGDSRLQLNFRAT 378
Db 281 SPSRXYFOYEGSLTPCAQGVITVFNQTVMLSAKQHLTLSDTLWPGDSRLQLNFRAT 340
QY 379 QPLNGRVIBASFPAGVDSSPRAAEFVQLNSCLAAGDILALVFGLLFAVTSVAFVQMRQ 438
Db 341 QPLNGRVIBASFPAGVDSSPRAAEFVQLNSCLAAGDILALVFGLLFAVTSVAFVQMRQ 400
QY 439 HRRGTGGSVYRPA 452
Db 401 HRRGTGGSVYRPS 414

```

Search completed: March 5, 2004, 09:08:29  
 Job time : 24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

Run on: March 5, 2004, 09:06:36 ; Search time 40 Seconds  
(without alignments)  
2422.984 Million cell updates/sec

```

Title:      US-09-967-237-2
Perfect score: 2424
Sequence:   1 MAPLCPFWLPLIIPAPAG.....RRGKGGVSYRPAEVAETGA 459

```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

```
Database :
1: /cgn2_6/ptodata/1/pubpaaf/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaaf/PCr_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaaf/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaaf/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaaf/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaaf/PCrT5_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaaf/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaaf/US09_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaaf/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaaf/US09C_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaaf/US09C_NEW_PUB.pcp.*
12: /cgn2_6/ptodata/1/pubpaaf/US10_PUBCOMB.pcp.*
13: /cgn2_6/ptodata/1/pubpaaf/US10C_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaaf/US10C_NEW_PUB.pcp.*
15: /cgn2_6/ptodata/1/pubpaaf/US10C_NEW_PUB.pcp.*
16: /cgn2_6/ptodata/1/pubpaaf/US06_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/1/pubpaaf/US06_PUBCOMB.pcp.*
18: /cgn2_6/ptodata/1/pubpaaf/US06_PUBCOMB.pcp.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2424	100.0	459	9	US-09-772-719-2	Sequence 2, Appl1
2	2424	100.0	458	10	US-09-967-237-2	Sequence 2, Appl1
3	2424	100.0	458	14	US-10-301-822-12	Sequence 12, Appl1
4	2424	100.0	459	15	US-10-463-572-10	Sequence 10, Appl1
5	2424	100.0	459	15	US-10-295-027-306	Sequence 106, Appl1
6	2424	100.0	458	15	US-10-295-027-1339	Sequence 1239, Appl1
7	2419	99.8	610	9	US-09-783-708-1	Sequence 1, Appl1
8	2007	92.8	377	10	US-09-967-237-87	Sequence 87, Appl1
9	1370	56.5	257	10	US-09-967-237-51	Sequence 51, Appl1
10	1364	56.3	256	9	US-09-772-719-51	Sequence 51, Appl1
11	884	36.5	170	9	US-09-772-719-54	Sequence 54, Appl1
12	562	23.2	337	10	US-09-946-374-423	Sequence 423, Appl1
13	562	23.2	337	10	US-09-983-000A-27	Sequence 27, Appl1
14	562	23.2	337	13	US-10-006-867-74	Sequence 74, Appl1
15	562	23.2	337	13	US-10-002-586-268	Sequence 268, Appl1

16	562	23.2	337	13	US-10-063-594-74	Sequence 268, App
17	562	23.2	337	14	US-10-174-559-268	Sequence 268, App
18	562	23.2	337	14	US-10-176-758-268	Sequence 268, App
19	562	23.2	337	14	US-10-175-73-268	Sequence 268, App
20	562	23.2	337	14	US-10-063-615-74	Sequence 74, App
21	562	23.2	337	14	US-10-173-708-268	Sequence 268, App
22	562	23.2	337	14	US-10-175-739-268	Sequence 268, App
23	562	23.2	337	14	US-10-175-755-268	Sequence 268, App
24	562	23.2	337	14	US-10-176-482-268	Sequence 268, App
25	562	23.2	337	14	US-10-176-757-268	Sequence 268, App
26	562	23.2	337	14	US-10-176-912-268	Sequence 268, App
27	562	23.2	337	14	US-10-180-552-268	Sequence 268, App
28	562	23.2	337	14	US-10-180-551-268	Sequence 268, App
29	562	23.2	337	14	US-10-063-502-74	Sequence 74, App
30	562	23.2	337	14	US-10-173-700-268	Sequence 268, App
31	562	23.2	337	14	US-10-174-572-268	Sequence 268, App
32	562	23.2	337	14	US-10-174-579-268	Sequence 268, App
33	562	23.2	337	14	US-10-174-582-268	Sequence 268, App
34	562	23.2	337	14	US-10-174-588-268	Sequence 268, App
35	562	23.2	337	14	US-10-175-739-268	Sequence 268, App
36	562	23.2	337	14	US-10-175-744-268	Sequence 268, App
37	562	23.2	337	14	US-10-175-743-268	Sequence 268, App
38	562	23.2	337	14	US-10-176-489-268	Sequence 268, App
39	562	23.2	337	14	US-10-176-492-268	Sequence 268, App
40	562	23.2	337	14	US-10-176-747-268	Sequence 268, App
41	562	23.2	337	14	US-10-176-750-268	Sequence 268, App
42	562	23.2	337	14	US-10-176-985-268	Sequence 268, App
43	562	23.2	337	14	US-10-176-987-268	Sequence 268, App
44	562	23.2	337	14	US-10-176-992-268	Sequence 268, App
45	562	23.2	337	14	US-10-176-993-268	Sequence 268, App

## ALIGNMENTS

RESULT 1.  
 US-09-772-719-2  
 Sequence 2, Application US/09772719  
 Patent No. US20020137910A1  
 GENERAL INFORMATION:  
 APPLICANT: Zavada, Jan  
 APPLICANT: Pastorekova, Silvia  
 APPLICANT: Pastorek, Jaromir  
 TITLE OF INVENTION: MN Gene and Protein  
 NUMBER OF SEQUENCES: 86  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Leona L. Lauder  
 STREET: 369 Pine Street  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/772,719  
 FILING DATE: 30-JAN-2001  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/485,049  
 FILING DATE: 07-JUN-1995  
 APPLICATION NUMBER: US 08/260,190  
 FILING DATE: 15-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lauder, Leona L.  
 REGISTRATION NUMBER: 30,863  
 REFERENCE/DOCKET NUMBER: D-0021.3E  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-981-2034  
 TELEFAX: 415-981-0332



INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 459 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: First 37 amino acids represent  
signal peptide, and remaining amino acids  
DESCRIPTION: represent mature protein  
US-09-772-719-2

Query Match 100.0%; Score 2424; DB 9; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1.6e-181;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSFWLPLIPAPAPGLTVQLLSILLMPVHPQRLPRMQEDSPFGGSSGSDPDL 60  
DB 1 MAPLCPSFWLPLIPAPAPGLTVQLLSILLMPVHPQRLPRMQEDSPFGGSSGSDPDL 60  
QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
QY 121 DPEPQNNARHDKKEDDQSHWRVYGGDPPMPRVSPACAGFQSPVDIRPOLAFCPLRLPL 180  
DB 121 DPEPQNNARHDKKEDDQSHWRVYGGDPPMPRVSPACAGFQSPVDIRPOLAFCPLRLPL 180  
QY 181 ELGFOPLPELRLANNHSHVOLTPPGLEMAIGPGEYRALQJLHMGAAGRGSEHT 240  
DB 181 ELGFOPLPELRLANNHSHVOLTPPGLEMAIGPGEYRALQJLHMGAAGRGSEHT 240  
QY 241 VEGHFPFAEIHVHLSFAFVDEALGRPGGLAVLAFLBEGPEENSAYEQLLSRLEBIA 300  
DB 241 VEGHFPFAEIHVHLSFAFVDEALGRPGGLAVLAFLBEGPEENSAYEQLLSRLEBIA 300  
QY 301 EGGSETOVPGDISALLPDSFRFYEGSLTTPPCAGQVITVFNQTVMLSAKQHLTIS 360  
DB 301 EGGSETOVPGDISALLPDSFRFYEGSLTTPPCAGQVITVFNQTVMLSAKQHLTIS 360  
QY 361 DTLMPGDSRLQNFRAQPLNGRVTEASFPAGVDSPPRAAPVOLNSCLAAGDILAVF 420  
DB 361 DTLMPGDSRLQNFRAQPLNGRVTEASFPAGVDSPPRAAPVOLNSCLAAGDILAVF 420  
QY 421 GLLFAVTSVAFVQMRQRHRRGTGGSYRPAEVAETGA 459  
DB 421 GLLFAVTSVAFVQMRQRHRRGTGGSYRPAEVAETGA 459

RESULT 2  
US-09-967-237-2  
Sequence 2, Application US/09967237  
Publication No. US20030049628A1  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
FILE REFERENCE: D-0021.5B-2  
CURRENT APPLICATION NUMBER: US/09/967,237  
CURRENT FILING DATE: 2001-09-27  
PRIOR APPLICATION NUMBER: 09/178,115  
PRIOR FILING DATE: 1998-10-23  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 2  
LENGTH: 459  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-967-237-2

Query Match 100.0%; Score 2424; DB 10; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1.6e-181;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSFWLPLIPAPAPGLTVQLLSILLMPVHPQRLPRMQEDSPFGGSSGSDPDL 60  
DB 1 MAPLCPSFWLPLIPAPAPGLTVQLLSILLMPVHPQRLPRMQEDSPFGGSSGSDPDL 60  
QY 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
DB 61 GEEDLPSEEDSPREEDPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
QY 121 DPEPQNNARHDKKEDDQSHWRVYGGDPPMPRVSPACAGFQSPVDIRPOLAFCPLRLPL 180  
DB 121 DPEPQNNARHDKKEDDQSHWRVYGGDPPMPRVSPACAGFQSPVDIRPOLAFCPLRLPL 180  
QY 181 ELGFOPLPELRLANNHSHVOLTPPGLEMAIGPGEYRALQJLHMGAAGRGSEHT 240  
DB 181 ELGFOPLPELRLANNHSHVOLTPPGLEMAIGPGEYRALQJLHMGAAGRGSEHT 240  
QY 241 VEGHFPFAEIHVHLSFAFVDEALGRPGGLAVLAFLBEGPEENSAYEQLLSRLEBIA 300  
DB 241 VEGHFPFAEIHVHLSFAFVDEALGRPGGLAVLAFLBEGPEENSAYEQLLSRLEBIA 300  
QY 301 EGGSETOVPGDISALLPDSFRFYEGSLTTPPCAGQVITVFNQTVMLSAKQHLTIS 360  
DB 301 EGGSETOVPGDISALLPDSFRFYEGSLTTPPCAGQVITVFNQTVMLSAKQHLTIS 360  
QY 361 DTLMPGDSRLQNFRAQPLNGRVTEASFPAGVDSPPRAAPVOLNSCLAAGDILAVF 420  
DB 361 DTLMPGDSRLQNFRAQPLNGRVTEASFPAGVDSPPRAAPVOLNSCLAAGDILAVF 420  
QY 421 GLLFAVTSVAFVQMRQRHRRGTGGSYRPAEVAETGA 459  
DB 421 GLLFAVTSVAFVQMRQRHRRGTGGSYRPAEVAETGA 459

RESULT 3  
US-10-301-822-12  
Sequence 12, Application US/10301822  
Publication No. US20030148410A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Berger, Allison  
APPLICANT: Guillemette, Tracy L.  
APPLICANT: Kamatkar, Shubangi  
APPLICANT: Schlegel, Robert  
APPLICANT: Monahan, John E.  
APPLICANT: Thibodeau, Stephen N.  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
FILE REFERENCE: PPM01-029P22NM  
CURRENT APPLICATION NUMBER: US/10/301,822  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US 60/339,971  
PRIOR FILING DATE: 2001-12-10  
PRIOR APPLICATION NUMBER: US 60/361,978  
PRIOR FILING DATE: 2002-03-05  
PRIOR APPLICATION NUMBER: US 60/381,988  
PRIOR FILING DATE: 2002-05-20  
NUMBER OF SEQ ID NOS: 228  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 459  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-301-822-12

Query Match 100.0%; Score 2424; DB 14; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1.6e-181;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSFWLPLIPAPAPGLTVQLLSILLMPVHPQRLPRMQEDSPFGGSSGSDPDL 60

Db	1	MAPLCEPFWLPLIIPAPRGLTYQLLSLLLMFHPQRLPMQDSDSLGGSGSGEDDPL	60
Qy	61	GEBDLPEEDSPREEDPPGEGEDLPGEGEDLPEYKPKSEEGSLKEDLPTVEAPG	120
Db	61	GEBDLPEEDSPREEDPPGEGEDLPGEGEDLPEYKPKSEEGSLKEDLPTVEAPG	120
Qy	121	DPOGPONNARHDEGDDQSHWRYGCPMPMPRVSPACAGFOSPVDIRQLAFCPALRPL	180
Db	121	DPOGPONNARHDEGDDQSHWRYGCPMPMPRVSPACAGFOSPVDIRQLAFCPALRPL	180
Qy	181	ELLGFLQPLPELRLRNNCHSVQTLTLPGLLENALPGREYRALQLHLHWGAAGRPGSEHT	240
Db	181	ELLGFLQPLPELRLRNNCHSVQTLTLPGLLENALPGREYRALQLHLHWGAAGRPGSEHT	240
Qy	241	VEGHRPAELHVHVLSTAFPAVDEALGRGGLAVLAALLEGSPENSNAYEQLSRLEEA	300
Db	241	VEGHRPAELHVHVLSTAFPAVDEALGRGGLAVLAALLEGSPENSNAYEQLSRLEEA	300
Qy	301	EEGSETOVPGDLIDISALLPSDFSRHYFOYEGSLTTPCAQGVIMTFNQTVMLSAQHLTIS	360
Db	301	EEGSETOVPGDLIDISALLPSDFSRHYFOYEGSLTTPCAQGVIMTFNQTVMLSAQHLTIS	360
Qy	361	DTLWGPEDSLQLNFRATQPLNGRYIEASFPAGVSSPRAPVQUNSCLAAGDILALVF	420
Db	361	DTLWGPEDSLQLNFRATQPLNGRYIEASFPAGVSSPRAPVQUNSCLAAGDILALVF	420
Qy	421	GLLFAVTSVAFIVQMRQRGRGKGVSRRPAVEAETA	459
Db	421	GLLFAVTSVAFIVQMRQRGRGKGVSRRPAVEAETA	459
RESULT 4			
US-10-465-572-10			
; Sequence 10, Application US/10465572			
; Publication No. US20030207840A1			
; GENERAL INFORMATION:			
APPLICANT: Rigging, Gregory			
APPLICANT: Lal, Anita			
TITLE OF INVENTION: GENES INDUCED BY HYPOXIA			
FILE REFERENCE: 000250.00012			
CURRENT APPLICATION NUMBER: US/10/465,572			
CURRENT FILING DATE: 2003-06-20			
PRIOR APPLICATION NUMBER: US/10/201,642			
PRIOR FILING DATE: 2002-07-24			
PRIOR APPLICATION NUMBER: 60/307,600			
PRIOR FILING DATE: 2001-07-26			
NUMBER OF SEQ ID NOS: 30			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 10			
LENGTH: 459			
Type: PRT			
ORGANISM: Homo sapiens			
US-10-465-572-10			

	Query Match	Similarity	100.0%	Score 2424	DB 15	Length 459
Best Local	Similarity	100.0%	Pred. No. 1	6e-181		
Matches	455	Conservative	0	Mismatches	0	Indels
					0	Gaps
					0	
QY	1	MAELCSPMPLIIPAPAGLT	TVOLLISLLIMVPHQRL	PRMOEDS	PLCGSSGGDDPL	60
Db	1	MAELCSPMPLIIPAPAGLT	TVOLLISLLIMVPHQRL	PRMOEDS	PLCGSSGGDDPL	60
QY	61	GSEEDLPSSEEDSPREEDP	PGSEEDLPSEEDLPSEEDLP	PEVKSSEEGSLKLEDD	PTVEAEG	120
Db	61	GSEEDLPSSEEDSPREEDP	PGSEEDLPSEEDLPSEEDLP	PEVKSSEEGSLKLEDD	PTVEAEG	120
QY	121	DPEPQNNNAHRDEGDDG	SHRKYGGDPPMPVSPACAG	RQSPVDIISPOLAAR	CCPLARPL	180
Db	121	DPEPQNNNAHRDEGDDG	SHRKYGGDPPMPVSPACAG	RQSPVDIISPOLAAR	CCPLARPL	180
QY	181	ELIGFOLPPLPELRANNG	SVQTLPRGLPMALGGR	ETRYALDHLHMGAA	RPSSEET	240

Db	181	ELLGFLPPLPELRNRNNGHSYQULTPPGLEMALGPGREYVAJOLHLMWGAAGPSEHT	243
Qy	241	VEGHRFPATIHVVHLSSTARVDEALGPPGLAVLAALFEGPEENSAYPEOLLSRLEEIA	3000
Db	241	VEGHRFPATIHVVHLSSTARVDEALGPPGLAVLAALFEGPEENSAYPEOLLSRLEEIA	3000
Qy	301	EEGSEFQVEGLDISALLPDPFSRYQYEGSLTTPFCAGGVMTVFNFQVWMLSAKOLHTLS	3660
Db	301	EEGSEFQVEGLDISALLPDPFSRYQYEGSLTTPFCAGGVMTVFNFQVWMLSAKOLHTLS	3660
Qy	361	DTIMGFGDSRLQLNFRATQPLNGRVIEASPPAGVDSBPRAAPVQNLSCLAAGDILALVF	4200
Db	361	DTIMGFGDSRLQLNFRATQPLNGRVIEASPPAGVDSBPRAAPVQNLSCLAAGDILALVF	4200
Qy	421	GLLFATVSVAFLVQMRQHRGRTKGAVYRPAEVAETGA	459
Db	421	GLLFATVSVAFLVQMRQHRGRTKGAVYRPAEVAETGA	459

```

RESULT 5
US-10-295-027-306
Sequence 306, US2003022350A1
Publication No. US2003022350A1
GENERAL INFORMATION:
APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevezi, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
FILE REFERENCE: 018501-012500US
CURRENT APPLICATION NUMBER: US/10/295,027
CURRENT FILING DATE: 2002-11-13
PRIORITY APPLICATION NUMBER: US 09/663,733
PRIORITY FILING DATE: 2000-09-15
PRIORITY APPLICATION NUMBER: US 60/350,666
PRIORITY FILING DATE: 2001-11-13
PRIORITY APPLICATION NUMBER: US 60/335,394
PRIORITY FILING DATE: 2001-11-15
PRIORITY APPLICATION NUMBER: US 60/332,464
PRIORITY FILING DATE: 2001-11-21
PRIORITY APPLICATION NUMBER: US 60/334,393
PRIORITY FILING DATE: 2001-11-29
PRIORITY APPLICATION NUMBER: US 60/340,376
PRIORITY FILING DATE: 2001-12-14
PRIORITY APPLICATION NUMBER: US 60/347,211
PRIORITY FILING DATE: 2002-01-08
PRIORITY APPLICATION NUMBER: US 60/347,349
PRIORITY FILING DATE: 2002-01-10
PRIORITY APPLICATION NUMBER: US 60/355,250
PRIORITY FILING DATE: 2002-02-08
PRIORITY APPLICATION NUMBER: US 60/356,714
PRIORITY FILING DATE: 2002-02-13
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1366
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 306
LENGTH: 459
TYPE: PRT
ORGANISM: Homo sapiens
US-10-295-027-306

```

Query Match 100.0%; Score 2424; DB 15; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1.6e-181;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

```

Db      1  MAPLCPSPWMLPLIAPAPGLTYQLLSLLMLPVHQPLPMQEDSPGCGSSGSDPDL 60
Qy      61  GGEDLPSEEDSPREDDPGSEDLPGSEDLPEVKKSEEGSLKLEDLPTVEARG 120
Db      61  GGEDLPSEEDSPREDDPGSEDLPGSEDLPEVKKSEEGSLKLEDLPTVEARG 120
Qy      121  DPOEPQNNAHKDEGDQSHMRVGGDPMPMPRVSPACAGFQSPVDIRPQLAACFCALRPL 180
Db      121  DPOEPQNNAHKDEGDQSHMRVGGDPMPMPRVSPACAGFQSPVDIRPQLAACFCALRPL 180
Qy      181  ELIGFQLPPLPELRIRNNHGSVOLTPPGLEMAIGREYRALQHLHMGAAGRPSEHT 240
Db      181  ELIGFQLPPLPELRIRNNHGSVOLTPPGLEMAIGREYRALQHLHMGAAGRPSEHT 240
Qy      241  VEGHRFPAPAEIHVHLSTAFARVDEALGRPGGLAVLAALFEEGPEENSAYEQLLSRLEBIA 300
Db      241  VEGHRFPAPAEIHVHLSTAFARVDEALGRPGGLAVLAALFEEGPEENSAYEQLLSRLEBIA 300
Qy      301  EGGSETQVPGDLSALPSPDSRYFOYEGSLTTPPCAQGVIMTFVNOTWMLSAKOLHTIS 360
Db      301  EGGSETQVPGDLSALPSPDSRYFOYEGSLTTPPCAQGVIMTFVNOTWMLSAKOLHTIS 360
Qy      361  DTLWGPDSRLQINFRATQPLNGRVIEASFPAGVDSPPRAAEVQNSCLAAGDILALVF 420
Db      361  DTLWGPDSRLQINFRATQPLNGRVIEASFPAGVDSPPRAAEVQNSCLAAGDILALVF 420
Qy      421  GLFPAVTSVAFVQMRQHRGTGKGVSYRPAEVAETGA 459
Db      421  GLFPAVTSVAFVQMRQHRGTGKGVSYRPAEVAETGA 459

RESULT 6
US-10-295-027-1239
/ Sequence 1239, Application US/10295027
/ Publication No. US2003023250A1
/ GENERAL INFORMATION:
/ APPLICANT: Afar, Daniel
/ APPLICANT: Aziz, Natsasha
/ APPLICANT: Ginsberg, Wendy M.
/ APPLICANT: Gish, Kurt C.
/ APPLICANT: Glynn, Richard
/ APPLICANT: Hevezi, Peter A.
/ APPLICANT: Mack, David H.
/ APPLICANT: Murray, Susan R.
/ APPLICANT: Watson, Richard
/ APPLICANT: Eos Biotechnology, Inc.
/ TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
/ TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
/ FILE REFERENCE: 018501-012500US
/ CURRENT APPLICATION NUMBER: US/10/295,027
/ CURRENT FILING DATE: 2002-11-13
/ PRIOR APPLICATION NUMBER: US 09/663,733
/ PRIOR FILING DATE: 2000-09-15
/ PRIOR APPLICATION NUMBER: US 60/350,666
/ PRIOR FILING DATE: 2001-11-13
/ PRIOR APPLICATION NUMBER: US 60/335,394
/ PRIOR FILING DATE: 2001-11-15
/ PRIOR APPLICATION NUMBER: US 60/332,464
/ PRIOR FILING DATE: 2001-11-21
/ PRIOR APPLICATION NUMBER: US 60/334,393
/ PRIOR FILING DATE: 2001-11-29
/ PRIOR APPLICATION NUMBER: US 60/340,376
/ PRIOR FILING DATE: 2001-12-14
/ PRIOR APPLICATION NUMBER: US 60/347,211
/ PRIOR FILING DATE: 2002-01-08
/ PRIOR APPLICATION NUMBER: US 60/347,349
/ PRIOR FILING DATE: 2002-01-10
/ PRIOR APPLICATION NUMBER: US 60/355,250
/ PRIOR FILING DATE: 2002-02-08
/ PRIOR APPLICATION NUMBER: US 60/356,714
/ PRIOR FILING DATE: 2002-02-13
/ Remaining Prior Application data removed - See File Wrapper or PALM.

```

```

/ NUMBER OF SEQ ID NOS: 1386
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 1239
/ LENGTH: 459
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-295-027-1239
Query Match      100.0%; Score 2424; DB 15; Length 459;
Best Local Similarity 100.0%; Pred. No. 1,6e-181;
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1  MAPLCPSPWMLPLIAPAPGLTYQLLSLLMLPVHQPLPMQEDSPGCGSSGSDPDL 60
Db      1  MAPLCPSPWMLPLIAPAPGLTYQLLSLLMLPVHQPLPMQEDSPGCGSSGSDPDL 60
Qy      61  GGEDLPSEEDSPREDDPGSEDLPGSEDLPEVKKSEEGSLKLEDLPTVEARG 120
Db      61  GGEDLPSEEDSPREDDPGSEDLPGSEDLPEVKKSEEGSLKLEDLPTVEARG 120
Qy      121  DPOEPQNNAHKDEGDQSHMRVGGDPMPMPRVSPACAGFQSPVDIRPQLAACFCALRPL 180
Db      121  DPOEPQNNAHKDEGDQSHMRVGGDPMPMPRVSPACAGFQSPVDIRPQLAACFCALRPL 180
Qy      181  ELIGFQLPPLPELRIRNNHGSVOLTPPGLEMAIGREYRALQHLHMGAAGRPSEHT 240
Db      181  ELIGFQLPPLPELRIRNNHGSVOLTPPGLEMAIGREYRALQHLHMGAAGRPSEHT 240
Qy      241  VEGHRFPAPAEIHVHLSTAFARVDEALGRPGGLAVLAALFEEGPEENSAYEQLLSRLEBIA 300
Db      241  VEGHRFPAPAEIHVHLSTAFARVDEALGRPGGLAVLAALFEEGPEENSAYEQLLSRLEBIA 300
Qy      301  EGGSETQVPGDLSALPSPDSRYFOYEGSLTTPPCAQGVIMTFVNOTWMLSAKOLHTIS 360
Db      301  EGGSETQVPGDLSALPSPDSRYFOYEGSLTTPPCAQGVIMTFVNOTWMLSAKOLHTIS 360
Qy      361  DTLWGPDSRLQINFRATQPLNGRVIEASFPAGVDSPPRAAEVQNSCLAAGDILALVF 420
Db      361  DTLWGPDSRLQINFRATQPLNGRVIEASFPAGVDSPPRAAEVQNSCLAAGDILALVF 420
Qy      421  GLFPAVTSVAFVQMRQHRGTGKGVSYRPAEVAETGA 459
Db      421  GLFPAVTSVAFVQMRQHRGTGKGVSYRPAEVAETGA 459

```

```

RESULT 7
US-09-783-708-1
/ Sequence 1, Application US/09783708
/ Patent No. US20020058041A1
/ GENERAL INFORMATION:
/ APPLICANT: The Regents of the University of California
/ TITLE OF INVENTION: KIDNEY-SPECIFIC TUMOR VACCINE DIRECTED AGAINST KIDNEY TUMOR ANT
/ FILE REFERENCE: 3067-897420US
/ CURRENT APPLICATION NUMBER: US/09/783,708
/ CURRENT FILING DATE: 2001-02-13
/ PRIOR APPLICATION NUMBER: 60/182,429
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: 60/182,636
/ PRIOR FILING DATE: 2000-02-15
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 1
/ LENGTH: 610
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: G250-GM-CSF fusion protein
US-09-783-708-1

```

```

Query Match      99.8%; Score 2419; DB 9; Length 610;
Best Local Similarity 100.0%; Pred. No. 5.6e-181;
Matches 458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 2 APLCPSPWLPPLIPAPAPGLTVQLLSLLLMPVHPQRLPRMOEDSPGSGSGEDDPLG 61
Db 147 APLCPSPWLPPLITAPAPGLTVQLLSLLLMPVHPQRLPRMOEDSPGSGSGEDDPLG 206
QY 62 EEDLPSEEDSPREDDPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 121
Db 207 EEDLPSEEDSPREDDPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 266
QY 122 PQEPQNNAHKDEDDQSHWRVYSGDPMPWPVSACAGRFQSPDIPRQALAFPCALRPE 181
Db 267 PQEPQNNAHKDEDDQSHWRVYSGDPMPWPVSACAGRFQSPDIPRQALAFPCALRPE 326
QY 182 LLGFGPLPLPLRLNNHGSVQLTLPPGLEMALGPGREYRALQHLHMGAGRPGESEHY 241
Db 327 LLGFGPLPLPLRLNNHGSVQLTLPPGLEMALGPGREYRALQHLHMGAGRPGESEHY 386
QY 242 EGHRRPAPATHVHSTAFARVDEALGRPGGLAVLAFLLEGPEENSAYEQLLSRLBEIAE 301
Db 387 EGHRRPAPATHVHSTAFARVDEALGRPGGLAVLAFLLEGPEENSAYEQLLSRLBEIAE 446
QY 302 EGSETQVPGDLISALPDSFRYFYQYEGSLTTPPCAQGVIMVFNQTMLSAKQHTLSD 361
Db 447 EGSETQVPGDLISALPDSFRYFYQYEGSLTTPPCAQGVIMVFNQTMLSAKQHTLSD 506
QY 362 TLWGDSRLQNLNFRATQPLNGRVIEASFPAGVDSPPRAEPVQNSCLAAGDITLAVFG 421
Db 507 TLWGDSRLQNLNFRATQPLNGRVIEASFPAGVDSPPRAEPVQNSCLAAGDITLAVFG 566
QY 422 LLFAVTSVAFVQWRQHRQRTKGVSYRPAEVAETGA 459
Db 567 LLFAVTSVAFVQWRQHRQRTKGVSYRPAEVAETGA 604

RESULT 8
US-09-967-237-87
; Sequence 87, Application US/09967237
; Publication No. US20030049828A1
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; FILE REFERENCE: D-0021.5B-2
; CURRENT APPLICATION NUMBER: US/09/967,237
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/178,115
; PRIOR FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
; LENGTH: 377
; TYPE: PRT
; ORGANISM: HUMAN
US-09-967-237-87

Query Match 82.8%; Score 2007; DB 10; Length 377;
Best Local Similarity 100.0%; Pred. No. 5,6e-149;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 181 REYRALQHLHMGAGRPGESEHYEGHRRPAPATHVHSTAFARVDEALGRPGGLAVLA 240
QY 278 FLEGPEENSAYEQLLSRLBEIAEGSEYQVPGDLISALPDSFRYFYQYEGSLTTPPCA 337
Db 241 FLEGPEENSAYEQLLSRLBEIAEGSEYQVPGDLISALPDSFRYFYQYEGSLTTPPCA 300
QY 338 QGVIMVFNQTMLSAKQHTLSDITLWGDSRLQNLNFRATQPLNGRVIEASFPAGVDS 397
Db 301 QGVIMVFNQTMLSAKQHTLSDITLWGDSRLQNLNFRATQPLNGRVIEASFPAGVDS 360
QY 398 PRAEPVQNSCLAAGD 414
Db 361 PRAEPVQNSCLAAGD 377

RESULT 9
US-09-967-237-51
; Sequence 51, Application US/09967237
; Publication No. US20030049828A1
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; FILE REFERENCE: D-0021.5B-2
; CURRENT APPLICATION NUMBER: US/09/967,237
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/178,115
; PRIOR FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51
; LENGTH: 257
; TYPE: PRT
; ORGANISM: HUMAN
US-09-967-237-51

Query Match 56.5%; Score 1370; DB 10; Length 257;
Best Local Similarity 100.0%; Pred. No. 2.7e-99;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/772,719  
FILING DATE: 30-JAN-2001  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 256 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: carbonic anhydrase domain  
US-09-772-719-51

Query Match 56.3%; Score 1364; DB 9; Length 256;  
Best Local Similarity 100.0%; Pred. No. 7.8e-99;  
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 DDQSHWRYGGDPMPRVSACAGRFQSPVDIRPOLAFCPALRPPELLGFOUPPLPELR 195  
DB 1 DDQSHWRYGGDPMPRVSACAGRFQSPVDIRPOLAFCPALRPPELLGFOUPPLPELR 60  
QY 196 RNNHGSVQLTLPGLLEALGPGREYRALQLHLWGAAAGRGSGHTVEGRFPALHVVHL 255  
DB 61 RNNHGSVQLTLPGLLEALGPGREYRALQLHLWGAAAGRGSGHTVEGRFPALHVVHL 120  
QY 256 STAFARVDEALGRPGGLAVLAAPLEEGPEENSAYEQILSRLEIEAEGSETOVPGLDISA 315  
DB 121 STAFARVDEALGRPGGLAVLAAPLEEGPEENSAYEQILSRLEIEAEGSETOVPGLDISA 180  
QY 316 LLPDFFSRFYFOYEGSLTTPPCAGVITWVFNQVWMSAKQLHTLSLTWNGPDSRLQLNF 375  
DB 181 LLPDFFSRFYFOYEGSLTTPPCAGVITWVFNQVWMSAKQLHTLSLTWNGPDSRLQLNF 240  
QY 376 RATOPPLNGRVIEASFP 391  
DB 241 RATOPPLNGRVIEASFP 256

RESULT 11  
US-09-772-719-54  
Sequence 54, Application US/09772719  
Patent No. US20020137910A1  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder

STREET: 369 Pine Street  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/772,719  
FILING DATE: 30-JAN-2001  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 54:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 170 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-772-719-54

Query Match 36.5%; Score 884; DB 9; Length 170;  
Best Local Similarity 100.0%; Pred. No. 1.8e-61;  
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 RALQLHLWGAAAGRGPGSEHTVEGRFPALHVVHLSTAFARVDEALGRPGGLAVLAAPLE 280  
DB 1 RALQLHLWGAAAGRGPGSEHTVEGRFPALHVVHLSTAFARVDEALGRPGGLAVLAAPLE 60  
QY 281 EGPENSAYEQILSRLEIEAEGSETOVPGLDISALPDSRIFYOYEGSLTTPPCAGV 340  
DB 61 EGPENSAYEQILSRLEIEAEGSETOVPGLDISALPDSRIFYOYEGSLTTPPCAGV 120  
QY 341 IWTVFNQVWMSAKQLHTLSLTWNGPDSRLQLNFRATOPPLNGRVIEASF 390  
DB 121 IWTVFNQVWMSAKQLHTLSLTWNGPDSRLQLNFRATOPPLNGRVIEASF 170

RESULT 12  
US-09-946-374-423  
Sequence 423, Application US/09946374  
Publication No. US20030073129A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Baton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurrey, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2830P1C1  
CURRENT APPLICATION NUMBER: US/09/946,374  
CURRENT FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: 60/098716  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098723  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098749  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098750  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098803  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/098821  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/098843  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/099536  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099586  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099598  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099602  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099642  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099741  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099754  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099763  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099792  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099808  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099812  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099815  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099816  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/100385  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100388  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100390  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100584  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100627  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100661  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100662  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100664  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100683  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100684  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100710  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100711  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100848  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/100849  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/100919  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100930  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/101014  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/101068  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/101071  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/101279  
PRIOR FILING DATE: 1998-09-22  
PRIOR APPLICATION NUMBER: 60/101471  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101472  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101474  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101475  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101476  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101477  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101479  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101738  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101741  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101743  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101915  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101916  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/102207  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102240  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102307  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102330  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102331  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102484  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 60/102487  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 60/102570  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 60/102571  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 60/102684  
PRIOR FILING DATE: 1998-10-01  
PRIOR APPLICATION NUMBER: 60/102687  
PRIOR FILING DATE: 1998-10-01  
PRIOR APPLICATION NUMBER: 60/102965  
PRIOR FILING DATE: 1998-10-02  
PRIOR APPLICATION NUMBER: 60/103258  
PRIOR FILING DATE: 1998-10-06  
PRIOR APPLICATION NUMBER: 60/103314  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103315  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103328  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103395

PRIOR FILING DATE: 1998-10-07  
 PRIOR APPLICATION NUMBER: 60/103396  
 PRIOR FILING DATE: 1998-10-07  
 PRIOR APPLICATION NUMBER: 60/103401  
 PRIOR FILING DATE: 1998-10-07  
 PRIOR APPLICATION NUMBER: 60/103449  
 PRIOR FILING DATE: 1998-10-06  
 PRIOR APPLICATION NUMBER: 60/103633  
 PRIOR FILING DATE: 1998-10-08  
 PRIOR APPLICATION NUMBER: 60/103678  
 PRIOR FILING DATE: 1998-10-08  
 PRIOR APPLICATION NUMBER: 60/103679  
 PRIOR FILING DATE: 1998-10-08  
 PRIOR APPLICATION NUMBER: 60/103711  
 PRIOR FILING DATE: 1998-10-08  
 PRIOR APPLICATION NUMBER: 60/104257  
 PRIOR FILING DATE: 1998-10-14  
 PRIOR APPLICATION NUMBER: 60/104987  
 PRIOR FILING DATE: 1998-10-20  
 PRIOR APPLICATION NUMBER: 60/105000  
 PRIOR FILING DATE: 1998-10-20  
 PRIOR APPLICATION NUMBER: 60/105002  
 PRIOR FILING DATE: 1998-10-20  
 PRIOR APPLICATION NUMBER: 60/105104  
 PRIOR FILING DATE: 1998-10-21  
 PRIOR APPLICATION NUMBER: 60/105169  
 PRIOR FILING DATE: 1998-10-22  
 PRIOR APPLICATION NUMBER: 60/105266  
 PRIOR FILING DATE: 1998-10-22  
 PRIOR APPLICATION NUMBER: 60/105693  
 PRIOR FILING DATE: 1998-10-26  
 PRIOR APPLICATION NUMBER: 60/105694  
 PRIOR FILING DATE: 1998-10-26  
 PRIOR APPLICATION NUMBER: 60/105807

Query Match 23.2%; Score 562; DB 10; Length 337;  
 Best Local Similarity 38.8%; Pred. No. 7,7e-36;  
 Matches 128; Conservative 47; Mismatches 129; Indels 26; Gaps 7;

QY 136 DDQSHWRY---GGDPMPRVSPACAGRFQSPVDIRPOLAACFPALPRLLELGFOLPPLPE 192  
 DB 17 DGGQHWTEGPHGQDHPASYPBCGNNASPPIDQTDVTFDPLPALQPHGVDQGTPEP 76  
 QY 193 LRLRNHGSVOLTPPLGEMALGPGRERYALQTLHMGAAGRP-GSEHTVEGHRFPALPIH 251  
 DB 77 LDHNNHGVTLSPSTLYIG-GLPRKYVAAQHLHMGQKSGSGSEHQINSEATFAELH 135  
 QY 252 VVHL-STAFARVDEALGRPGGLAVLAFLIEGSEENSAYBQLLSRLIEEISSETOYFG 310  
 DB 136 IVHYDSDSYDSLSEAAERPOGLAVLGILIEVGETKNIAVEHILSHLHVRHKDKTSVPP 195  
 QY 311 LDISALLPDSRFGYEGSLTPPCAGQVITVFNQVTWLSAKQLHTLSDTLWGPDSR 370  
 DB 196 FNRLBELPKQLGQYFRYNSLTTPCYQSVLTVFRRSQISMQLHQLTSTEEPP 255  
 QY 371 LQI---NFRATQPLNGRVIETASPPAGVDSPPRAAPVQNSCLAAGDIALVFGLLFA-- 425  
 DB 256 SKLVQNVYRALQPLNQHWVFASF-----IQAGSYTTGEMLSLGVGLVGL 302  
 QY 426 --VTSVAFIVQMRGRHRTKGGVSYRPAE 453  
 DB 303 CLLAVYFIARKIRKRLKRLKRNKSVVFTSAQ 332

# RESULT 13

US-09-983-000A-27  
 Sequence 27, Application US/09983000A  
 Publication No. US20030118585A1  
 GENERAL INFORMATION:  
 APPLICANT: AGY Therapeutics  
 APPLICANT: Melcher, Thorsten  
 APPLICANT: Mueller, Sabine  
 APPLICANT: Chin, Daniel

TITLE OF INVENTION: USE OF PROTEIN BIOMOLECULAR TARGETS IN THE TREATMENT AND VISTO  
 TITLE OF INVENTION: OF BRAIN TUMORS  
 FILE REFERENCE: 263/180 -- Peagleman -- AGY  
 CURRENT APPLICATION NUMBER: US/09/983,000A  
 CURRENT FILING DATE: 2001-10-17  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 27  
 LENGTH: 337  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: Gene  
 LOCATION: (1)..(337)  
 OTHER INFORMATION: Carbonic anhydrase domain of human carbonic anhydrase VIX  
 US-09-983-000A-27

Query Match 23.2%; Score 562; DB 10; Length 337;  
 Best Local Similarity 38.8%; Pred. No. 7,7e-36;  
 Matches 128; Conservative 47; Mismatches 129; Indels 26; Gaps 7;

QY 136 DDQSHWRY---GGDPMPRVSPACAGRFQSPVDIRPOLAACFPALPRLLELGFOLPPLPE 192  
 DB 17 DGGQHWTEGPHGQDHPASYPBCGNNASPPIDQTDVTFDPLPALQPHGVDQGTPEP 76  
 QY 193 LRLRNHGSVOLTPPLGEMALGPGRERYALQTLHMGAAGRP-GSEHTVEGHRFPALPIH 251  
 DB 77 LDHNNHGVTLSPSTLYIG-GLPRKYVAAQHLHMGQKSGSGSEHQINSEATFAELH 135  
 QY 252 VVHL-STAFARVDEALGRPGGLAVLAFLIEGSEENSAYBQLLSRLIEEISSETOYFG 310  
 DB 136 IVHYDSDSYDSLSEAAERPOGLAVLGILIEVGETKNIAVEHILSHLHVRHKDKTSVPP 195  
 QY 311 LDISALLPDSRFGYEGSLTPPCAGQVITVFNQVTWLSAKQLHTLSDTLWGPDSR 370  
 DB 196 FNRLBELPKQLGQYFRYNSLTTPCYQSVLTVFRRSQISMQLHQLTSTEEPP 255  
 QY 371 LQI---NFRATQPLNGRVIETASPPAGVDSPPRAAPVQNSCLAAGDIALVFGLLFA-- 425  
 DB 256 SKLVQNVYRALQPLNQHWVFASF-----IQAGSYTTGEMLSLGVGLVGL 302  
 QY 426 --VTSVAFIVQMRGRHRTKGGVSYRPAE 453  
 DB 303 CLLAVYFIARKIRKRLKRLKRNKSVVFTSAQ 332

# RESULT 14

US-10-006-867-74  
 Sequence 74, Application US/10006867  
 Publication No. US20020119130A1  
 GENERAL INFORMATION:  
 APPLICANT: Eaton, Dan L.  
 APPLICANT: Filvaroff, Ellen  
 APPLICANT: Gerritsen, Mary B.  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Godowski, Paul J.  
 APPLICANT: Grimaldi, Christopher J.  
 APPLICANT: Gurney, Auecin L.  
 APPLICANT: Watanabe, Colin K.  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 FILE REFERENCE: P3330R1C1  
 CURRENT APPLICATION NUMBER: US/10/006,867  
 CURRENT FILING DATE: 2001-12-06  
 PRIOR APPLICATION NUMBER: 60/063435  
 PRIOR FILING DATE: 1997-10-29  
 PRIOR APPLICATION NUMBER: 60/064215  
 PRIOR FILING DATE: 1997-10-29  
 PRIOR APPLICATION NUMBER: 60/082797  
 PRIOR FILING DATE: 1998-04-22  
 PRIOR APPLICATION NUMBER: 60/083495  
 PRIOR FILING DATE: 1998-04-29



[illegible]

PRIOR APPLICATION NUMBER: 60/120014  
PRIOR FILING DATE: 1999-02-11  
PRIOR APPLICATION NUMBER: 60/129122  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: 60/129674  
PRIOR FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/131291  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/138387  
PRIOR FILING DATE: 1999-06-09  
PRIOR APPLICATION NUMBER: 60/144791  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/163495  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: 60/175481  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: 60/191007  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/199397  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/380139  
PRIOR FILING DATE: 1998-08-25  
PRIOR APPLICATION NUMBER: 09/311832  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 09/380137  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/380138  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/380142

## Query Match

23.2%; Score 562; DB 13; Length 337;

Best Local Similarity 38.8%; Pred. No. 7, 7e-36;  
Matches 128; Conservative 47; Mismatches 129; Indels 26; Gaps 7;

QY 136 DDGSHRY--GDBPWRVSPACAGRFQSPVDRPOLAFCCALPRLLELGPQPLPE 192  
DB 17 DGGQHTYEGPHGQDHWASYPBCGNNASPIDQDTSVTFDDLPALOPHGVDGCTEP 76  
QY 193 LRLRNGHVSQVLTLPPIELMALGPGREYRALQHLHWAAGR-PGSEHTVEGHRPAEIH 251  
DB 77 LDHNNHGTQLSJPSSTLYLG-GLPRKYVAQAHLHFWGQSGSGSHQINSBATPAEIH 135  
QY 252 VVHL-STAFARVDLHGPGGLAVLAFLERGPENAVYQULSRLEIEBSSFPQVPG 310  
DB 136 IVHYDSYDLSSEAEPRPGLAVLGLIEVGETKNIAVEHILSHLEVHKQKXSVPP 195  
QY 311 LDISALPDSFSRYEQYEGSLTPPCAQGVITWFENQTVMLSAKQULTLSDTLWGPDSR 370  
DB 196 FHLRELPLKQJGQYRVNGSLTTPCYQSVLMTVFYRSGQISMEQLEKLGTLFSTBEP 255  
QY 371 LQJ--NFRATOPPLNGRYTEASFPAGVDSSFRAAEPVQULNSCLAAGDILALVFGGLFA-- 425  
DB 256 SKLLVQNYRALQPLQKRVFASF-----IQAGSSYTTGEMSLGVGILVGC 302  
QY 426 --VTSAPFLVQMRQRGRGTGQSVYRPAE 453  
DB 303 CILLAVYFIARKIRKRIENRKSXVYFISAQ 332

RESULT 15  
US-10-052-586-268  
Sequence 268, Application US/10052586  
Publication No. US20020127584A1  
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C1  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: US/10/052,586  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059266  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063120  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063121  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063486  
PRIOR FILING DATE: 1997-10-21  
PRIOR APPLICATION NUMBER: 60/063540  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063541  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063544  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063564  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063734  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063870  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/064103  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066120  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066466  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066772  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069335  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069425  
PRIOR FILING DATE: 1997-12-12  
PRIOR APPLICATION NUMBER: 60/069870  
PRIOR FILING DATE: 1997-12-17  
PRIOR APPLICATION NUMBER: 60/068017  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/07450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/07632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/07886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08

PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086023  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/086392  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/086486  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087098  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087208  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087609  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087759  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087827  
PRIOR FILING DATE: 1998-06-03  
PRIOR APPLICATION NUMBER: 60/088025  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088028  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088029  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088033  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088167  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088202  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088212

PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088217  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088326  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088655  
PRIOR FILING DATE: 1998-06-09  
PRIOR APPLICATION NUMBER: 60/088722  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088738  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088740  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088811  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088824  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088825  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088826  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088861  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088863  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088876  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/089090  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089105  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089512  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089514  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089598  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089908

## Query Match

23.2%; Score 562; DB 13; Length 337;

Best Local Similarity 38.8%; Pred. No. 7; 7e-36;

Matches 128; Conservative 47; Mismatches 129; Indels 26; Gaps 7;

QY 136 DDQSHWRY---GDBPWVRVSPACAGRFQSPVDIRPQLAFCPALRLEHLGQPLPE 192  
DB 17 DGGQWTEGPHGQDHWPAASYPCGNAOSTIDQDTSVTFDPDLPALOPHGIDQPTET 76  
QY 193 LRLRNNGHVSQVLTTPGJENALGPGREYRALQTLHFWGAGRP-GSEHTVEGHRFPAEIH 251  
DB 77 LDHNNGHVQLSLPSTLYLG-GLPRKYVAAQHLHFWGQKSGPSGSEHQINSEATFAELH 135  
QY 252 VYHL-STPAFYDEALGPGGLATVLAFLBEPBENAYQQLSRLEIEBESSTQVPG 310  
DB 136 IYHYDSYDYDSLSEAEHPQGLAVLGLIEVETKXNAYHILSHLEVHKQKTSVPP 195  
QY 311 LDISALPSDFSRVQYEGSLTTPCAQGVMTVFQNTWLSAKQTLTSLDTITGPGDSR 370  
DB 196 FULRELLPKQLOQYRYNGSLTTPCYQSVLMTVFYRSGIWSHQLEKQGLPSTEEEP 255  
QY 371 LQI---NFRATQPLNGRYTEASFPAQYDSSFPRAEPVQUNSCLAAGIILVGLFLA-- 425  
DB 256 SKLIVQNYRALQPLNQRAWFASF-----IQAGSSYTTGMLSLGVGILVGL 302  
QY 426 --VTSAPFVQWRROHRRGTGKGVSYRPAE 453  
DB 303 CULAVYFIARKIRKRLKRNKSVFTPSAQ 332

Search completed: March 5, 2004, 09:13:52

Mon Mar 8 10:16:31 2004

us-09-967-237-2.rapb

Page 12

Job time : 41 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 09:00:14 ; Search time 18 Seconds

(without alignments)  
1327.789 Million cell updates/sec

Title: US-09-967-237-2

Perfect score: 2424  
Sequence: 1 MAPLCPSPWLPFLIPAPAPG.....RRGTGVSYPRAVAETGA 459

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2424	100.0	459	CAH9_HUMAN	Q16790 homo sapien
2	1596	65.8	437	CAH9_MOUSE	Q8VBD5 mus musculu
3	562	23.2	337	CAH9_HUMAN	Q9ULX7 homo sapien
4	558	23.0	337	CAH9_MOUSE	Q9WTF6 mus musculu
5	537.5	22.2	354	CAH9_HUMAN	Q43570 homo sapien
6	526	21.7	307	CAH9_SHEEP	P08600 ovis aries
7	512	21.1	355	CAH9_RABBIT	P08630 ovis aries
8	511	21.1	319	CAH9_BOVIN	P18915 bos taurus
9	496	20.5	354	CAH9_MOUSE	Q8C185 mus musculu
10	488	20.1	308	CAH9_HUMAN	P23280 homo sapien
11	467.5	19.3	317	CAH9_MOUSE	P18761 mus musculu
12	422.5	17.4	324	CAH9_MOUSE	Q99N23 mus musculu
13	422	17.4	259	CAH9_TRIKX	Q8UW45 tribolodon
14	419.5	17.3	290	CAH9_MOUSE	P28651 mus musculu
15	412	17.0	289	CAH9_HUMAN	P35219 homo sapien
16	409	16.9	264	CAH9_HUMAN	P43166 homo sapien
17	402.5	16.6	260	CAH9_BRAE	Q92051 brachydanto
18	398.5	16.4	255	CAH9_MOUSE	Q9EQD8 mus musculu
19	391	16.1	259	CAH9_BOVIN	P00919 bos taurus
20	379.5	15.7	259	CAH9_HUMAN	P00918 homo sapien
21	379.5	15.7	259	CAH9_MOUSE	P00920 mus musculu
22	379	15.6	259	CAH9_SHEEP	P00922 ovis aries
23	377	15.6	262	CAH9_MOUSE	Q96611 mus musculu
24	376	15.5	262	CAH9_HUMAN	Q8N1Q1 homo sapien
25	370	15.3	260	CAH9_MOUSE	P13334 mus musculu
26	368.5	15.2	259	CAH9_RABBIT	P00919 ovicollagus
27	368.5	15.2	259	CAH9_RAT	P27139 rattus norv
28	366.5	15.1	260	CAH9_MACMU	P00916 macaca mura
29	364.5	15.0	259	CAH9_CHICK	P35217 macaca neme
30	362.5	15.0	260	CAH9_MACNE	P35217 macaca neme
31	359.5	14.8	260	CAH9_HUMAN	F00915 homo sapien
32	354	14.6	260	CAH9_HORSE	F00917 equus cabal
33	353.5	14.6	259	CAH9_RAT	P14141 rattus norv

RESULT 1	CAH9_HUMAN	STANDARD	PRT	459 AA.
ID	CAH9_HUMAN			
AC	Q16750			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Carbonic anhydrase IX precursor (BC 4.2.1.1) (carbonate dehydratase IX) (CA-IX) (CAIX) (Membrane antigen MN) (P54/58N) (Renal cell carcinoma-associated antigen G250) (RCC-associated antigen G250)			
DE	(PMR1).			
GN	CA9 OR MN OR G250.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RC	TISSUE=Carcinoma;			
RX	MEDLINE=9436734; PubMed=8084592;			
RA	Pastorek X., Pastorekova S., Callebaut I., Mornon J.-P., Zelink V., Oavsky R., Zet, Ovicova M., Liao S., Portetelle D., Stanbridge E.J., Zavada J., Burny A., Kettmann R.,			
RA	"Cloning and characterization of MN, a human tumor-associated protein with a domain homologous to carbonic anhydrase and a putative helix-loop-helix DNA binding segment.";			
RT	Oncogene 9:2877-2888(1994).			
RL	[2]			
RN	SEQUENCE FROM N.A., AND VARIANT VAL-33.			
RP	TISSUE=Renal cell carcinoma;			
RX	MEDLINE=20175484; PubMed=10709109;			
RA	Grabmaier K., Vissers J.L.M., De Weijert M.C.A., Oosterwijk-Wakka U.C., Van Bokhoven A., Brakenhoff R.H., Noessner E., Mulders P.A., Merks G., Figdor C.G., Adema G.J., Oosterwijk E.,			
RA	"Molecular cloning and immunogenicity of renal cell carcinoma-associated antigen G250.";			
RT	Int. J. Cancer 85:865-870(2000).			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	TISSUE=Colon;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Rata S.S., Loquellano N.A., Peters G.J., Abtman R.D., Mullaly S.J., Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Wotley K.C., Hale S., Garcia A.M., Gay L.J., Huliyil S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.N.,			

## ALIGNMENTS

34	352	14.5	262	1	CAH1_MONDO	Q8HY33 monodelphis
35	352	14.5	265	1	CAH4_MOUSE	Q64444 mus musculu
36	350.5	14.5	309	1	CAH3_MOUSE	P16015 mus musculu
37	347	14.3	309	1	CAH4_RAT	P48284 rattus norv
38	342.5	14.1	259	1	CAH3_HUMAN	P07451 homo sapien
39	342	14.1	304	1	CAH5_RAT	P43165 rattus norv
40	341	14.1	260	1	CAH1_SHEEP	P48282 ovis aries
41	340.5	14.0	317	1	CAH5_HUMAN	Q9Y260 homo sapien
42	339.5	14.0	317	1	CAH5_MOUSE	Q9G260 mus musculu
43	338.5	14.0	259	1	CAH3_HORSE	P07450 equus cabal
44	334	13.8	309	1	CAH5_CAEEL	Q10462 caenorhabdi
45	333	13.7	308	1	CAH4_RABBIT	P48283 ovicollagus

RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schercher A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN (4)  
 RN CHARACTERIZATION.  
 RX MEDLINE=92188536; PubMed=1312272;  
 RX Pastorekova S., Zavadova Z., Kostal M., Babusikova O., Zavadova J.;  
 RT "A novel quasi-viral agent, Matu, is a two-component system.";  
 RL Virology 187:620-626(1992).  
 RN (5)  
 RN SUBCELLULAR LOCATION.  
 RX MEDLINE=93252505; PubMed=8486430;  
 RA Zavadova J., Zavadova Z., Pastorekova S., Ciampor F., Pastorek J.,  
 RA Zelnik V.;  
 RT "Expression of Matu-MN protein in human tumor cultures and in clinical  
 RT specimens.";  
 RL Int. J. Cancer 54:268-274(1993).  
 CC -1- FUNCTION: Reversible hydration of carbon dioxide. May be involved  
 CC in the control of cell proliferation and transformation. Appears  
 CC to be a novel specific biomarker for a cervical neoplasia.  
 CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.  
 CC -1- COFACTOR: Zinc (By similarity).  
 CC -1- SUBUNIT: Forms oligomers linked by disulfide bonds.  
 CC -1- SUBCELLULAR LOCATION: FOUND ON THE SURFACE MICROVILLI AND IN THE  
 CC NUCLEUS, PARTICULARLY IN NUCLEOLUS.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN CARCINOMA CELLS LINES.  
 CC EXPRESSION IS RESTRICTED TO VERY FEW NORMAL TISSUES AND THE MOST  
 CC ABUNDANT EXPRESSION IS FOUND IN THE EPITHELIAL CELLS OF GASTRIC  
 CC MUCOSA.  
 CC -1- PTM: N-glycosylated.  
 CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isdb-eb.ch/announce/>  
 CC or send an email to [license@isdb-eb.ch](mailto:license@isdb-eb.ch)).  
 CC -----  
 CC EMBL: X66839; CAA47315.1; -  
 CC EMBL: AJ010586; CAB82444.1; -  
 CC EMBL: BC014950; AAH14950.1; -  
 CC PIR: I38013; I38013.  
 CC HSR: P00918; I38013.  
 CC Gene: HGN:1383; CA9.  
 CC MIM: 603179; -  
 CC GO: GO:0016021; C:integral to membrane; TAS.  
 CC GO: GO:0004089; F:carbonate dehydratase activity; TAS.  
 CC InterPro: IPR001148; Euk Combd.  
 CC Pfam: PF00194; carb\_anhydrase; 1.  
 CC PRODOM: PD000865; Euk\_Combd; 1.  
 CC PROSITE: PS00162; Euk\_CO2\_ANHYDRASE; 1.  
 CC Lysase; Zinc; Transmembrane; Glycoprotein; Antigen; Signal;  
 KW Nuclear protein; Polymorphism.  
 FT SIGNAL 1 37  
 FT CHAIN 38 459  
 FT METAL 226 226  
 FT METAL 228 228  
 FT METAL 231 231  
 FT TRANSMEM 415 435  
 FT CARBOHYD 346 346  
 FT VARIANT 33 33  
 FT /FTID=VAR 010787.  
 FT M -> V.  
 SQ SEQUENCE 459 AA; 49729 MW; 5E3C1E2936050B3F CRC64;  
 Query Match 100.0%; Score 2424; DB 1; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 5,9e-130;  
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLCPSPWPLPLIPAPAPGLTVOLLSTLLMPVAPORLP3NQEDSP1GGSSGDEDEL 60  
 DB 1 MAPLCPSPWPLPLIPAPAPGLTVOLLSTLLMPVAPORLP3NQEDSP1GGSSGDEDEL 60  
 QY 61 GEEPLPEEDSPREDEPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 DB 61 GEEPLPEEDSPREDEPPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP 120  
 QY 121 DPQEPQNNARHDEKGDQSHMRYGSDPMPWRVSPACAGFQSGFVDIRPOLAFCPALRPL 180  
 DB 121 DPQEPQNNARHDEKGDQSHMRYGSDPMPWRVSPACAGFQSGFVDIRPOLAFCPALRPL 180  
 QY 181 ELLEFQPLPELPLRNNGHSVOLLTPPGLEMAAGREYRALQTLHGAAGRPSSEPT 240  
 DB 181 ELLEFQPLPELPLRNNGHSVOLLTPPGLEMAAGREYRALQTLHGAAGRPSSEPT 240  
 QY 241 VEGHFPFAEIHVHLSAPFAVDEALGRPGGLAVLAFAEEGEEENSAVEQLSLREETA 300  
 DB 241 VEGHFPFAEIHVHLSAPFAVDEALGRPGGLAVLAFAEEGEEENSAVEQLSLREETA 300  
 QY 301 EEGSETQVPGDLISALLPSDFSRFYQYEGSLTPPCAGVITVFNQVTMLSAKQILHTLS 360  
 DB 301 EEGSETQVPGDLISALLPSDFSRFYQYEGSLTPPCAGVITVFNQVTMLSAKQILHTLS 360  
 QY 361 DTLMGPDSDRLQNFRAOTPLNGRVLEASPPAGVDSPPAEPVQNSCLAAGDITAIYF 420  
 DB 361 DTLMGPDSDRLQNFRAOTPLNGRVLEASPPAGVDSPPAEPVQNSCLAAGDITAIYF 420  
 QY 421 GLFPAVTSVAFVQMRORHRTGKGVSYPAPAEVETGA 459  
 DB 421 GLFPAVTSVAFVQMRORHRTGKGVSYPAPAEVETGA 459  
 QY 421 GLFPAVTSVAFVQMRORHRTGKGVSYPAPAEVETGA 459  
 DB 421 GLFPAVTSVAFVQMRORHRTGKGVSYPAPAEVETGA 459  
 RESULT 2  
 CAH9\_MOUSE STANDARD; PRT; 437 AA.  
 ID CAH9\_MOUSE  
 AC Q8VHB5; Q8VHB5; Q8VDE4;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Carbonic anhydrase IX precursor (EC 4.2.1.1) (Carbonate dehydratase IX) (CA-IX) (CAIX) (Membrane antigen MN homolog).  
 GN CA9 OR CAH9.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=129/Ola;  
 RX PubMed=14604546;  
 RA Zaitovicova M., Tarabkova K., Syatova E., Gibadulinova A., Mucha V., Jakubikova L., Blesova Z., Rafajova M., Ortova Gut M.O., Parkkila S., Parkkila A.-K., Mahed A., Sly W.S., Horak I., Pastorek J., Pastorekova S.;  
 RA "Monoclonal antibodies generated in carbonic anhydrase IX-deficient mice recognize different domains of tumour-associated hypoxia-induced carbonic anhydrase IX.";  
 RL J. Immunol. Methods 282:117-134(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC Ortova M.;  
 RX Submitted (DEC-1999) to the EMBL/GenBank/DBD databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=ICR; TISSUE=Small intestine;  
 RA Wang Y.P., Yoshikawa K., Kozaki K., Miyaihi O., Nakagawa A., Murakatsu H., Kawada Y., Uchida K., Nishikawa N., Saga S.;  
 RA "Alternative spliced mRNA coding for MN/CA9.";  
 RT Submitted (JUN-2002) to the EMBL/GenBank/DBD databases.  
 CC -1- FUNCTION: Reversible hydration of carbon dioxide (By similarity).  
 CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.  
 CC -1- COFACTOR: Zinc (By similarity).

```

CC -1- SUBUNIT: Forms oligomers linked by disulfide bonds (By
CC similarity)
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8VHB5-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8VHB5-2; Sequence=VSP_007409, VSP_007410;
CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; AY049077; AAL14193.1; -
DR EMBL; AJ245857; CAC80975.1; -
DR EMBL; AB086322; BAC00816.1; -
DR MGD; MGI:2447188; Car9.
DR InterPro; IPR001148; Euk_Coanhd.
DR Pfam; PF00194; carb_anhydase; 1.
DR PRODOM; PD000865; Euk_Coanhd; 1.
DR PROSITE; PS00162; Euk_CO2_ANHYDRASE; 1.
KW Lyase; Zinc; Transmembrane; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 31
FT CHAIN 32 437 CARBONIC ANHYDRASE IX.
FT METAL 205 205 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 207 207 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 230 230 ZINC (CATALYTIC) (BY SIMILARITY).
FT TRANSMEM 391 411 POTENTIAL.
FT CARBOHYD 325 325 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPIC 282 282 G -> V (in isoform 2).
FT VARSPIC 283 437 /FTId=VSP_007409.
FT VARSPIC 437 /FTId=VSP_007410.
SQ SEQUENCE 437 AA; 47264 MW; 88F23380DCD3344 CRC64;
Query Match 65.8%; Score 1596; DB 1; Length 437;
Best Local Similarity 69.4%; Pred. No. 2, 6e-83;
Matches 320; Conservative 36; Mismatches 79; Indels 26; Gaps 6;
QY 1 MAPLCSPMLPLIPAPAGLTVALLLSLMLPVHPOQLPRMGSDFLGSSGGEDEPL 60
DB 1 MASLGSPWAPLSTPAP---TAQLLLFLLQLQVSAQPOGLSGMGEPSELGSSSGEDS-L 55
QY 61 GEEDLPSEEDSPREEDPPEBEDLPGEEDLPGEEDLPVKPKSEEGSKLDELPTVEAPG 120
DB 56 GUDVLPSEEDAPPEADP-----PGEEDPPEVNSERNMESLGLDLSTPERP- 102
QY 121 DPQEPONNAHRDQGDQSHRWYGGDPWFVRVSPACAGRFQSPVDIRPOLAFCPALRPL 180
DB 103 ---EHSQSHGDEKGGSHSWYSGTLLMPQVSPACAGRFQSPVDIRLERTAFQRTLOPL 159
QY 181 ELIGPOLPLPRLAKLNNGHSVQLTLPGLKMALEPGEYRLQLHLHWGAAGPGESEHT 240
DB 160 ELIGYELQPLPELSTLNNGHTVQLTLPGLKMALEPGEYRLQLHLHWGSDHPGSHT 219
QY 241 VEGHRRPAEIHVHLSSTARVDEALGRGGLAVLAALFEGPEENSAVEQLLRLEBIA 300
DB 220 VNGHRRPAEIHVHLSSTARVDEALGRGGLAVLAALFQEPENSAVEQLLRLEBIS 279
QY 301 EESSETOVPGCLDLSALLPDSFRYROYEGSLTTPPCAGQVMTVNCVMSAQCLHTLS 360
DB 280 EESGSKIEIPGLDVALLPDSLDRYYRSGSLTTPPCSGQVMTVNEVTKSIAQLHTLS 339
QY 361 DLTMGQDSRLQINFATQPLNGRVLEAFSPAGVSSPPRAEPVQVNLNCLAGDILALVF 420
DB 340 VSLMGTRDSRLQINFATQPLNGRTLEAFSPAAEDSSP---EPVHVNCSCTFAGDILALVF 396

```

```

QY 421 GILFAVTSVAFINQMRQHR--GTKGVSYPAEVAETGA 459
DB 397 GILFAVTSVAFINQMRQHRSGTKGVSYPAEVAETGA 437
RESULT 3
CAHE HUMAN STANDARD; PRT; 337 AA.
ID CAHE HUMAN
AC 991X7; GANCF4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Carbonic anhydrase XIV precursor (EC 4.2.1.1) (Carbonate dehydratase
DE XIV) (CA-XIV).
GN CA14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=99443874; PubMed=10512682;
RA Fujikawa-Akachi K., Nishimori I., Taguchi T., Onishi S.;
RT "Human carbonic anhydrase XIV (CA14): cDNA cloning, mRNA expression,
RT and mapping to chromosome 1."
RL Genomics 61:74-81(1999).
RN [2]
RP SEQUENCE FROM N.A.
RP Isoagi T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto C., Makamatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuno Y., Ono T., Okano K., Toshiyaka Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NBD human cDNA sequencing project."
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RP TISSUE=Brain, Lung, and Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Haefl F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.J.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunatane P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A.C., Shevchenko Y., Bouffard G.G.,
RA Whiting M., Madan A., Young A., Green E.D., Dickson M.C.,
RA Blakesley R.W., Touchman D.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Schercher A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Reversible hydration of carbon dioxide.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -1- COFACTOR: Zinc (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: HIGH EXPRESSION IN ALL PARTS OF THE CENTRAL
CC NERVOUS SYSTEM AND LOWER EXPRESSION IN ADULT LIVER, HEART, SMALL
CC INTESTINE, COLON, KIDNEY, URINARY BLADDER, AND SKELETAL MUSCLE.
CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AB025904; BAA85002.1; -  
 DR EMBL: AC074765; BAC1191.1; -  
 DR EMBL: BC034412; AAH34412.1; -  
 DR HSSP: O43570; LUD0.  
 DR GeneW: HGNC:1372; CA14.  
 DR MIM: 604832; -  
 DR GO: GO:0016021; C:integral to membrane; TAS.  
 DR GO: GO:0004089; F:carbonate dehydratase activity; TAS.  
 DR InterPro: IPR001148; Euk\_Coanhd.  
 DR Pfam: PF00194; carb\_anhydase; 1.  
 DR ProDom: PD000865; Euk\_Coanhd; 1.  
 DR ProSITE: PS00162; Euk\_Co2\_ANHYDRASE; FALSE\_NEG.  
 KW Lyase; Zinc; Transmembrane; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 337  
 FT DOMAIN 16 290  
 FT TRANSMEM 291 311  
 FT DOMAIN 312 337  
 FT METAL 109 109  
 FT METAL 111 111  
 FT METAL 135 135  
 FT CARBOHYD 213 213  
 FT CONFLICT 229 229  
 FT SEQUENCE 337 AA; 37667 MW; 6E101C44BA70A700 CRC64;

Query Match 23.2%; Score 562; DB 1; Length 337;  
 Best Local Similarity 38.8%; Pred. No. 3,7e-25;  
 Matches 128; Conservative 47; Mismatches 129; Indels 26; Gaps 7;

QY 136 DDGSHWRY---GSPRPWRVSPACAGRFQSPVDRPQLAFCALPRLFLGPGQPLPBE 192  
 Db 17 DGGHWTYEGPHGQDHMPASYPGCGNNAOSFIDIDTSTVTFDDLPALOGHYDPTGTP 76  
 QY 193 LRLRNHGSVOLTPPGLEMAAGREYRALQLHLHWAAGRP-GSEHTVEGHRPAETH 251  
 Db 77 LDLNNGHTVOLSLPSTLYLG-GLPKRYVAQLHLHWAQSGSGSEHQLNSBATAELH 135  
 QY 252 VVHL-STAFARVDLHPPGSLAVLAFLLEGPENNAVYQLSLRELAESSEQVVG 310  
 Db 136 IVHYDSYDLSLSEAEKPPGLAVLGLILEVGETKNAYHILSHLHRAHKQKTSVFP 195  
 QY 311 LDISALPDSFRRFYQYEGSLTTPPCAGVYVFNQTVMLSAKQHLTSLDITLWPGDSR 370  
 Db 196 FNLRELPLKQGGYFRYNGSLTTPPCQSVLTVFYRRSGISWEQLEKLGTLFSTEEBP 255  
 QY 371 LQL---NFRATQPLNGRITENSFPAGVDSFRAEPVQNSCLAAGDILAVFGLLFA-- 425  
 Db 256 SKLTVQNYRALQPNQWRVFAF-----IQAGSYTTGEMSLVGLVGL 302  
 QY 426 -VTSVAFVQMERQHRGRGTGVSYPBAE 453  
 Db 303 CULLAVYFIARKIRKRIENKRSVVFISAQ 332

RESULT 4  
 CAHE MOUSE STANDARD; PRT; 337 AA.  
 ID CAHE MOUSE STANDARD; PRT; 337 AA.  
 AC Q9WV76;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 15-MAR-2004 (Rel. 43; Last annotation update)  
 DE Carbonic anhydrase XIV precursor (EC 4.2.1.1) (Carbonate dehydratase XIV) (CA-XIV).  
 GN CA14 OR CA14 OR CATV.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=99269110; PubMed=10336468;  
 RA Mori K., Ogawa Y., Ebihara K., Tamura N., Tashiro K., Kuwahara T., Mkiyama M., Sugawara A., Ozaki S., Tanaka I., Nakao K.;  
 RT "Isolation and characterization of CA XIV, a novel membrane-bound carbonic anhydrase from mouse kidney.";  
 RL J. Biol. Chem. 274:15701-15705(1999).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Stachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldi M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedl T.B., Toshiyuki S., Carninci P., Prange C., Rana S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Motley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W., Villalón D.K., Muzny D.M., Sodergren E.U., Lu X., Gibbs R.A., Faney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Rodriguez Y.S.N., Krzywicki M.J., Skalska U., Smalins D.E., Schnerch A., Schein J.E., Jones S.J.M., Maira M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Reversible hydration of carbon dioxide.  
 CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.  
 CC -1- COFACTOR: Zinc (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -1- TISSUE SPECIFICITY: Most abundant in the kidney and heart, followed by the skeletal muscle, brain, lung and liver.  
 CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AB005450; BAA78709.1; -  
 DR EMBL: BC046995; AAH46995.1; -  
 DR HSSP: O43570; LUD0.  
 DR WGI: WGI:1344341; Car14.  
 DR InterPro: IPR001148; Euk\_Coanhd.  
 DR Pfam: PF00194; carb\_anhydase; 1.  
 DR ProDom: PD000865; Euk\_Coanhd; 1.  
 DR ProSITE: PS00162; Euk\_Co2\_ANHYDRASE; FALSE\_NEG.

FT SIGNAL 1 15  
 FT CHAIN 16 337  
 FT DOMAIN 16 290  
 FT TRANSMEM 291 311  
 FT DOMAIN 312 337  
 FT METAL 109 109  
 FT METAL 111 111  
 FT METAL 135 135  
 FT CARBOHYD 213 213  
 FT SEQUENCE 337 AA; 37505 MW; 32F02F4DB78AC0C9 CRC64;

Query Match 23.0%; Score 558; DB 1; Length 337;  
 Best Local Similarity 39.4%; Pred. No. 6,2e-25;  
 Matches 130; Conservative 42; Mismatches 130; Indels 28; Gaps 8;

QY 136 DQSHWRY---GADPMPRPVSPACAGRFQSPVDRPOLAFCPALRPLELLGQPLPLE 132  
 Db 17 DGHHTVYTGPHGQCHMPHSYPCGGDQSPINTICTSDVITDPDPAVQPHGIDQTER 76  
 QY 193 LILRNNGSVQVLTLPPLGEMALGPREYRALQHLHWGAAGR-FGSHTWEGHFRPAEIH 251  
 Db 77 LHLNNGHTVQLSTPPTLHLG-GLPRKYTAQLHLHWQQRSLGSHQINSEVTAELH 135  
 QY 252 VVHL-STAFARVDEALGRPGGLAVLAFLGEGPRENSAVQELLRLLEIMEGSEGVPG 310  
 Db 136 VHHYDSQSYSSLSFNAQKPGGLAVGLILEGFTENPAYDILRLHEIRKCKQKTSVPP 195  
 QY 311 LDISALLPDSFRYQYEGSLTTPPCAGVITWVNTMLSAQOL----HTLSDTLMGP 366  
 Db 196 FSVRELFPOQLQFRRYNGSLTTPPCQSVLMTVNRRAQISMGQLEKLGTTLSSTEDP 255  
 QY 367 GDSRLQNFRAQVQINGRVIETASFAGVDSBPRAEPVQNSCLAAGDILALVGLLFA- 425  
 Db 256 SEPLVQ-NYRPVQPLNORTIFASF-----IQAGPLYTGBMLGLGVGILAGC 301  
 QY 426 ---VTSVAFVQMRQRRGRGTGKGVSYEPA 452  
 Db 302 LCLLAVYFLAQKIRKRLGNKSVVFTSA 331

## RESULT 5

CAHC\_HUMAN STANDARD; PRT; 354 AA.  
 AC 043570; Q9BWG2;  
 DT 15-JUL-1999 (rel. 38, created)  
 DT 15-JUL-1999 (rel. 38, last sequence update)  
 DT 10-OCT-2003 (rel. 42, last annotation update)  
 DE carbonic anhydrase XII precursor (EC 4.2.1.1) (Carbonate dehydratase  
 XII) (CA-XII) (Tumor antigen HOM-RCC-3.1.3).  
 GN CA12.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.  
 RC TISSUE=Renal cell carcinoma;  
 RX MEDLINE=98301622; PubMed=9636197;  
 RA Parkkila A.-K., Sahin U., Vollmar E., Siemer S., Goettfert E., Seitz G.,  
 RT Tuerci O., Sahin U., Vollmar E., Siemer S., Pfeundschnh M., Sly W.S.;  
 RT "Human carbonic anhydrase XII: cDNA cloning, expression, and  
 RT chromosomal localization of a carbonic anhydrase gene that is  
 RT overexpressed in some renal cell cancers.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:7608-7613(1998).  
 RN (2)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Lung;  
 RX MEDLINE=98445416; PubMed=9770531;  
 RA Ivanov S.V., Kuzmin I., Wei M.-H., Pack S., Geil L., Johnson B.E.,  
 RT Standridge E.J., Lerman M.I.;  
 RT "Down-regulation of transmembrane carbonic anhydrases in renal cell  
 RT carcinoma cell lines by wild-type von Hippel-Lindau transgenes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12596-12601(1998).  
 RN (3)  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Eye, and Kidney;  
 RX MEDLINE=2388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong J.,  
 RA Scapleton W., Soares W.B., Bonaldo W.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein W.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullicy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Hulik S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmitt J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN (4)  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 30-292.  
 RX MEDLINE=21396545; PubMed=11493685;  
 RA Whittington D.A., Maheo A., Ulmasov B., Shah G.N., Grubb J.H.,  
 RA Sly W.S., Christianson D.W.;  
 RT "Crystal structure of the dimeric extracellular domain of human  
 RT carbonic anhydrase XII: a bionic membrane protein overexpressed in  
 RT certain cancer tumor cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9545-9550(2001).  
 CC -1- FUNCTION: Reversible hydration of carbon dioxide.  
 CC -1- CATALYTIC ACTIVITY: H(2)O + CO(2) = H(2)O + H(2)O.  
 CC -1- COFACTOR: Zinc (By similarity).  
 CC -1- ENZYME REGULATION: Inhibited by acetazolamide.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q43570-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q43570-2; Sequence=VSP\_000772;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN COLON, KIDNEY, PROSTATE,  
 CC -1- INTESTINE AND ACTIVATED LYMPHOCYTES. EXPRESSED AT MUCH HIGHER  
 CC LEVELS IN THE RENAL CELL CANCERS THAN IN SURROUNDING NORMAL KIDNEY  
 CC TISSUE. MODERATELY EXPRESSED IN PANCREAS, OVARY, AND TESTIS.  
 CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, AF051882; AAC39789.1; -  
 CC EMBL, AF057335; AAC63952.1; -  
 CC EMBL, BC000278; AAH00278.1; -  
 CC EMBL, BC011691; AAH11691.1; -  
 CC EMBL, BC023981; AAH23981.1; -  
 CC PDB, 1JDO; 17-AUG-01.  
 CC PDB, 1JCZ; 17-AUG-01.  
 CC GeneW, HSMC:1371; CA12.  
 CC MIM, 603263; -  
 CC GO, GO:0016021; C:integral to membrane; TAS.  
 CC GO, GO:0004089; F:carbonate dehydratase activity; TAS.  
 CC GO, GO:0008270; F:zinc ion binding; TAS.  
 CC InterPro, IPR001148; Euk Coanhd.  
 CC Pfam, PF00194; carb anhydrase; 1.  
 CC ProDom, PD000865; Euk\_Coanhd; 1.  
 CC PROSITE, PS00162; EUK\_CO2\_ANGHYDRASE; 1.  
 CC KW Lyase; Zinc; Transmembrane; Signal; Alternative splicing;  
 CC 3D-structure.  
 CC SIGNAL 1 24  
 CC CHAIN 25 354  
 CC DOMAIN 25 301  
 CC TRANSMEM 302 322  
 CC DOMAIN 323 354  
 CC METAL 119 119  
 CC METAL 121 121  
 CC METAL 145 145  
 CC DISULFID 50 230  
 CC POTENTIAL.  
 CC CARBONIC ANHYDRASE XII.  
 CC EXTRACELLULAR (POTENTIAL).  
 CC POTENTIAL.  
 CC CYTOPLASMIC (POTENTIAL).  
 CC ZINC (CATALYTIC).  
 CC ZINC (CATALYTIC).  
 CC ZINC (CATALYTIC).  
 CC ZINC (CATALYTIC).

```

FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT VARSPLIC 292 302 Missing (in isoform 2).
/ftid=vsp_000772.

FT TURN 37 38
FT HELIX 40 42
FT HELIX 43 46
FT HELIX 48 51
FT STRAND 59 60
FT HELIX 62 64
FT STRAND 65 67
FT TURN 69 70
FT STRAND 75 77
FT STRAND 80 80
FT TURN 83 84
FT STRAND 86 91
FT STRAND 96 99
FT TURN 102 103
FT STRAND 105 107
FT STRAND 113 122
FT STRAND 124 124
FT TURN 125 126
FT STRAND 127 127
FT STRAND 134 135
FT TURN 136 137
FT STRAND 138 138
FT STRAND 142 150
FT TURN 151 153
FT HELIX 157 160
FT TURN 161 162
FT TURN 164 165
FT STRAND 167 176
FT HELIX 181 187
FT TURN 188 189
FT HELIX 190 193
FT TURN 196 197
FT STRAND 199 203
FT STRAND 206 206
FT HELIX 207 210
FT TURN 215 216
FT STRAND 218 223
FT TURN 228 229
FT STRAND 234 239
FT STRAND 243 245
FT HELIX 247 255
FT STRAND 258 258
FT TURN 262 263
FT STRAND 269 269
FT STRAND 285 287
SQ SEQUENCE 354 AA; 39451 MW; 9016216BF2CA6C0C CRC64;

Query Match 22.2%; Score 537.5; DB 1; Length 354;
Best Local Similarity 35.8%; Pred. No. 9.4e-24;
Matches 129; Conservative 46; Mismatches 142; Indels 43; Gaps 10;

117 EAPGDPPEPNNARDEKGDQSHRY---GDPFWRPVSPACAGRFQSPVDIRPOLAAF 173
19 EGPSP-APVNG-----SKMTYFGPDGNSWKXKPPSCGGLQSPIDLDHSDILQY 67
114 CPALPRLLELGFQRLPRLBELRLKNGHSVQLTLPGLLENALSGREYRALQHLHMGANG 233
68 DASLPLLEFGVYNSANKQFLITNNGHSVKLNLPSCDHQIQ-GLQSRYSATQHLHMGANG 126
234 RP-GSEHTVEGHRFPALIHVHL-STAFARVDEALGRPGGLAVLAFLBEGPENSAYEQ 291
127 DPHGSEHTVSGQHFAELHIVHNSDLYPDASTSKNSBGLAVLAVLIMG-SFNPSTYX 185
232 LLSLEIEIAEGSETQVPGDLISALLPDSRSRTQYEGSLTTPPCAGQGIWYFNQVWL 351
186 IFSHQHYKYGQGAFAVFGFIEELBERTAEYVRGSLTTPPCNPVTLMVTFNRPVQI 245
352 SAKOHLTSLDILM-----GPDGSRQLNFRATQPLNGRVIEASFAGVDSPPRAEPVQL 406

```

```

DB 246 SQEOLALLETALYCTHMDPSPREMINNRQYQKDERLVTYSF-----QV 292
QY 407 NSCLACDIALVYGLLFA-----VTSVAFVQMRQHRRTGKGVSYNPAEVAETGA 459
DB 293 QVCTAAGLSIGLILSLALAGILGICIVVVSIMLFRKSKKDKNGVATKATMETEA 352

RESULT 6
CAH6_SHEEP STANDARD; PRT; 307 AA.
AC P08060;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Carbonic anhydrase VI (EC 4.2.1.1) (Carbonate dehydratase VI) (CA-VI)
DE (Secreted carbonic anhydrase) (Salivary carbonic anhydrase).
GN CA6.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RC TISSUE=Saliva;
RX MEDLINE=88294021; PubMed=315834;
RA Fernley R.T., Wright R.D., Coghlan J.P.;
RT "Complete amino acid sequence of ovine salivary carbonic anhydrase.";
RL Biochemistry 27:2815-2820(1986).
CC -1- FUNCTION: Reversible hydration of carbon dioxide. Its role in
CC saliva is unknown.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -1- COFACTOR: Zinc (by similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Major constituent of saliva.
CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
CC family.
DR PIR: A29993; A29993.
DR HSP; O43570; IUD0.
DR InterPro: IPR001148; Euk Coanhd.
DR Pfam: PF00194; carb anhydrase; 1.
DR ProDom: PD000865; Euk Coanhd; 1.
DR PROSITE: PS00162; Euk CO2 ANHYDRASE; 1.
KW lysase; zinc; glycoprotein.
FT METAL 94 94 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 96 96 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 121 121 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 25 207 POTENTIAL.
FT CARBOHYD 50 50 N-LINKED (GLCNAC. . .).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .).
FT VARIANT 63 63 V -> M.
FT VARIANT 297 297 I -> M.
SQ SEQUENCE 307 AA; 35555 MW; 3368022D45B5D6C CRC64;

Query Match 21.7%; Score 526; DB 1; Length 307;
Best Local Similarity 43.2%; Pred. No. 3.6e-23;
Matches 111; Conservative 40; Mismatches 98; Indels 8; Gaps 4;

141 WRG-----GDPFWRPVSPACAGRFQSPVDIRPOLAFCPALPRLLELGFQRLPRLBELRLK 196
6 WTSSEGMDEAMHPLRFPYCGGRKSPIDLOKKYQVNPSPALNLTGYGLMH-GEPPVT 64
197 NNGHSVQLTLPGLLENALSGREYRALQHLHMGANG-R-PSSEHTVEGHRFPALIHVH 254
65 NNGHTVQISLPSMTSWTSDGTQYLAKQHFHMGASSEISGSEHTVGMAYIEIHVH 124
255 LSTAFARVDEALGRPGGLAVLAFLBEGPENSAYEQVPGDI 313
125 YNSKNSYEDAKQEPGGLAVLALVEVDYENAYYSKFISLBDIRYAGOSTVLRGLDI 184
314 SALLPSPDFRYPQYEGSLTTPPCAGQGIWYFNQVWLSAQLHTLSPTLWGPDSRLQL 373

```

DB 185 EDMLPGDLYVYYSYLSTTPCTENKMFVADIVYKLSKVQVEKLENSLHMQKTIQN 244

QY 374 NFRATQPLNGRVLEASF 390

DB 245 DYRRITQPLNHRVVEANF 261

RESULT 7

CAH6\_RABIT STANDARD; PRT; 355 AA.

AC Q6M230;

DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Carbonic anhydrase XII precursor (EC 4.2.1.1) (Carbonate dehydratase XII) (CA-XII).

GN CA12.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI\_Taxid=9986;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney cortex;

RX MEDLINE=22417091; PubMed=12388401;

RA Schwartz G.J., Kittelberger A.M., Watkins R.H., O'Reilly M.A.;

RT "Carbonic anhydrase XII mRNA encodes a hydrtatase that is differentially expressed along the rabbit nephron."

RL Am. J. Physiol. 264:P399-F410(2003).

CC -1- FUNCTION: Reversible hydration of carbon dioxide (By similarity).

CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.

CC -1- COFACTOR: Zinc (By similarity).

CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).

CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC

DR EMBL: AF263367; AAF91392.2; -

DR HSSP: O43570; JUDO.

DR Interpro: IPR001148; Euk Coanhd.

DR Pfam: PF00194; carb anhydrase; 1.

DR Prodom: PD000865; Euk Coanhd; 1.

DR PROSITE: PS00162; Euk\_CO2\_ANHYDRASE; 1.

KW Lyase; Zinc; Transmembrane; Signal.

FT SIGNAL 1 24

FT DOMAIN 25 355

FT DOMAIN 25 303

FT TRANSSEM 304 324

FT DOMAIN 325 355

FT METAL 119 119

FT METAL 121 121

FT METAL 145 145

FT DISULFID 50 230

FT CARBOHYD 28 28

FT CARBOHYD 80 80

FT CARBOHYD 88 88

FT CARBOHYD 162 162

SO SEQUENCE 355 AA; 58468580BBA437D5 CRC64;

QY 106 GSUKLBD--LPTVEAPDGPQEPONNAHRDKEDQDSHWRY---CGDPPWPRVPACAGR 159

DB 4 GSURAAAVLILGIAQUPSSAAPRNG-----SKRTYVGADGSEMSWKXKPPSCGR 53

Query Match 21.1%; Score 512; DB 1; Length 355;

Best Local Similarity 34.0%; Pred. No. 2,6e-22;

Matches 127; Conservative 58; Mismatches 146; Indels 42; Gaps 11;

QY 160 FQSPVDIRPQLAAPCPALRPIELIGFQIPLPELRLENNNGHSVOLITPGLAEMALGGRG 219

DB 54 LQSPIDILHGDTLQVDASLTPLFEGGYNVSADKQNLINDGSRVLNLPDMYLOGPSR- 112

QY 220 YEAQLQHLHWGAAGR-CESEHTEGHRFPABEIVVHL-STAFARVDEALGRPGGLAVIAA 277

DB 113 YTRQLHLHWGNRNDPYSEHHTVGKQPAALMELHIVHNSDSYPISTASKSSEGLAVIAV 172

QY 278 FLEEGPEENSAYEQLLSPLBEIABGSETVPGGLDISALLPDSFRRFYQEGSLTPPCA 337

DB 173 LIEKG-YENPYDRIFSFILRYVKYGVNVRIPFNIEELLPEKPAEYRYRGSITTPCY 231

QY 338 QGVITVENQYTMASAKQLHTLSDTLW-----GPGDSRLQNPFRATQPLNARVIBASFP 392

DB 232 PSVLTWTRNPTYSIQEQLMLQVALFTTRADDPAPPEMINNPPICNFGVRLVHASF-- 289

QY 393 GVDSSPRAAEVQUL--NSCLAAGDIAL---VEGLLFAVTSVAFLVOMERRQHRRTXG 446

DB 290 -----QEVQLPFGTGLSVGIILSVALAGVIGICIVLAVCIWFRFRKSKSGDNKG 340

QY 447 VSRPAEVAETGA 459

DB 341 VLNPAIMETBA 353

RESULT 8

CAH6\_BOVIN STANDARD; PRT; 319 AA.

ID CAH6\_BOVIN

AC P18915; Q95322; 16. Created

DT 01-NOV-1990 (Rel. 16, Created)

DT 15-OCT-1999 (Rel. 38, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Carbonic anhydrase VI precursor (EC 4.2.1.1) (Carbonate dehydratase VI) (CA-VI) (Secreted carbonic anhydrase) (Salivary carbonic anhydrase).

GN CA6.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

OX NCBI\_Taxid=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Submaxillary gland;

RX MEDLINE=96358528; PubMed=8761494;

RA Jiang W., Wolbach J.T., Gupta D.;

RT "Sequence of bovine carbonic anhydrase VI: potential recognition sites for N-acetylgalactosaminyltransferase."

RL Biochem. J. 318:291-296(1996).

RN [2]

R2 SEQUENCE OF 15-39.

R3 MEDLINE=89246331; PubMed=2497732;

RA Fernley R.T., Darling P., Aldred P., Wright R.D., Coghan J.P.;

RT "Tissue and species distribution of the secreted carbonic anhydrase isoenzyme."

RL Biochem. J. 259:91-96(1989).

CC -1- FUNCTION: Reversible hydration of carbon dioxide. Its role in saliva is unknown.

CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.

CC -1- COFACTOR: Zinc (By similarity).

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- TISSUE SPECIFICITY: Major constituent of saliva.

CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

```

CC -----
DR EMBL: X96503; CA65357.1; ...
DR PIR: S71877; S71877.
DR HSRP: Q43570; IJ00.
DR InterPro: IPR001148; Euk_Coanhd.
DR Pfam: PF00194; carb_anhydase; 1.
DR ProDom: PD000865; Euk_Coanhd; 1.
DR ProSITE: PS00162; Euk_CO2_ANGHYDRASE; 1.
DR Lyase; Zinc; Glycoprotein; Signal.
FT SIGNAL 1 14
FT CHAIN 1 319
FT METAL 106 319
FT METAL 106 319
FT METAL 108 319
FT METAL 133 319
FT DISULFID 37 219
FT CARBOHYD 62 62
FT CARBOHYD 251 251
FT CONFICT 16 16
SQ SEQUENCE 319 AA; 37006 MW; 17984A7A9083AED CRC64;

Query Match 21.1%; Score 511; DB 1; Length 319;
Best Local Similarity 42.0%; Pred. NO. 2.6e-22;
Matches 108; Conservative 40; Mismatches 97; Indels 12; Gaps 4;

QY 137 DQSHWRYGDDPPWPRVSPACRPFQSPVDIRPOLAARFPAIRPFIELGQPLPPIELRLR 196
DB 26 DEKHWRL-----QYDPCGGTROSPIDLKKKKRVNPSRLNLTGYGLRQ-GEPPMT 76
QY 197 NNGHSYQTLTPPGLEMAALGPGRYALQLHLHWG--AAGRPGSHTEGHRPAEIHVH 254
DB 77 NNGHTYQILSPSSMNTTSDSQYAKGMHPHMGDSSEISGSHYDGMKRIIEIHVH 136
QY 255 LSTAFRRVDALGRPGGLAVALFLE-EGPEBSAYEQLSLRELEIEEGSEYQVPELDI 313
DB 137 YHSKYGSEAEQNEEDGLAVALVEVDVYENTYNSFIHEDIRYAGQSTYLRDLDI 196
QY 314 SALLPSPDFRYQYEGSLTPPCAGVITWVENCYVWLSAQQLHTLSDTLWPGDSRLQL 373
DB 197 QDMLPDDLAVYYSYLSGLTTPSCNTENHMFVADTVKLSKQIETKLSNLTNOMENTON 256
QY 374 NFRATQPLNGRYIEAS 390
DB 257 NYRSTQPLNHRVVEANF 273

RESULT 9
CAHC_MOUSE STANDARD; PRT; 354 AA.
ID CAHC_MOUSE
AC Q8C185; Q8C185;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Carbonic anhydrase XII precursor (EC 4.2.1.1) (Carbonate dehydratase
DE XII) (CA_XII).
GN CA12 OR CAR12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=2234683; PubMed=1246851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nakazaki I., Osato N., Saito R., Suzuki H., Yamataka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bull C.P., Hume D.A., Quackenbush J.,
RA Schirini L.M., Kanapin A., Matsumura H., Baralov S., Beisel K.W.,
RA Blake U.A., Brand D., Brusic V., Chochina C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Komagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagasima T., Numata K., Okido T., Pavan M.J., Pereira G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Ring B.Z., Ringold M.,
RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawat U., Aizawa K., Aikawa T., Fukuda S.,
RA Hara A., Hashizume W., Imachi K., Ishi Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino K., Matselton R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Czech II;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.J., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang D., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Strappele M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Tohiyuki S., Carninci P., Prange C.,
RA Rabb S.S., Loguelfano N.A., Peters G.U., Abramson R.D., Malhy S.J.,
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faley J., Helton E., Kettman K., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalka U., Smaluk D.E.,
RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16703(2002).
CC - FUNCTION: Reversible hydration of carbon dioxide.
CC - CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC - COFACTOR: Zinc (By similarity).
CC - SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC - SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
CC family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AK052639; BAC35074.1;
CC EMBL: BC035941; AA035941.1;
CC InterPro: IPR001148; Euk_Coanhd.
CC Pfam: PF00194; carb_anhydase; 1.
CC ProDom: PD000865; Euk_Coanhd; 1.
CC ProSITE: PS00162; Euk_CO2_ANGHYDRASE; UNKNOWN_1.
CC Lyase; Zinc; Transmembrane; Signal.
FT SIGNAL 1 24
FT CHAIN 25 354
FT DOMAIN 25 301
FT TRANSMEM 302 322
FT DOMAIN 323 354
FT METAL 120 320
FT METAL 122 120
FT METAL 146 146

```

```

FT DISULFID 50 221 BY SIMILARITY.
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 297 297 D -> N (IN REF. 1).
SQ SEQUENCE 354 AA; 39695 MW; 98ED581BF2A16111 CRC64;

Query Match
Best Local Similarity 35.5%; Score 496; DB 1; Length 354;
Matches 115; Conservative 54; Mismatches 137; Indels 18; Gaps 7;

QY 139 SHWRY---GDPMPRVSPACAGRFQSPVDIRPOLAFCPALRPELLGQLPPLPELR 195
DB 30 SKWTYVGPAGEKXWMSKYPSCGGLQSPIDHSIDLQYDASIALPQGVVWSYEKLNLT 89
QY 196 RNNGHSVQULTPEGLMALGPGRERVALQLHLHNGAAGR--GSHHTYEGHFRPAELHYH 254
DB 90 TNDGSHVRLNLSMDYITQLOPHRYRAQLHLHNGNRNDPDSGSHTVSGHFAELHYH 149
QY 255 L-STARVYDEALGRPGGLAVLAFLBEGPENSAYEQLSLREIEAEGSETOVPLDI 313
DB 150 YNSDLYPDSTSDSKSEGLAVLAFLIEIG--SAPSYDKIFSHLGHVXKQOVLIPQFNI 208
QY 314 SALLPSDSRRYFQYSGSLTTPPCAGGVWTFVFNCTWMLSAKQHTLSLDTLM-----GPGD 368
DB 209 EELLPSPEYRYRYSGLTTPPCYPTVMTVRNPVQISQELALITALYFTHMDDPTP 268
QY 369 SRLQNFRAQPLNGRVLEASFPAGVDSPPAAEPVQINSLAAGDILALVFGLLPAVTS 428
DB 269 REMINFPQVQKFDRLVYISFRQGLLDTLSGLITSLVALA-----VLGIIST-VLA 321
QY 429 VAFIVQMRQHRRTKGVSTYRPA 452
DB 322 VSIWLFKRKKSKGDKGVYKPA 345

RESULT 10
CAH6_HUMAN STANDARD; PRT; 308 AA.
ID CAH6_HUMAN STANDARD; PRT; 308 AA.
AC P23280; Q96QX8; Q9UF03;
DT 01-NOV-1991 (Rel. 20; Created)
DT 28-FEB-2003 (Rel. 41; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Carbonic anhydrase VI precursor (EC 4.2.1.1) (Carbonate dehydratase
VI) (CA-VI) (secreted carbonic anhydrase) (salivary carbonic
anhydrase).
DE CA6.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91105141; PubMed=1899030;
RA Alfred P., Fu P., Barrett G., Penschow J.D., Wright R.,
RA Coghlan J.P., Fernley R.T.;
RT "Human secreted carbonic anhydrase: cDNA cloning, nucleotide
RT sequence, and hybridization histochemistry.";
RL Biochemistry 30:569-575(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Grubb D.J.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Frackland J.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Reversible hydration of carbon dioxide. Its role in
CC saliva is unknown.
CC -!- CATALYTIC ACTIVITY: H(2)O + CO(2) = HCO(3) + H(2)O.
CC -!- COFACTOR: Zinc (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.

```

```

CC -!- TISSUE SPECIFICITY: Major constituent of saliva.
CC -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M57892; AAF51892.1; -
DR EMBL; AF128418; AAF22565.1; JOINED.
DR EMBL; AF128411; AAF22565.1; JOINED.
DR EMBL; AF128412; AAF22565.1; JOINED.
DR EMBL; AF128413; AAF22565.1; JOINED.
DR EMBL; AF128414; AAF22565.1; JOINED.
DR EMBL; AF128415; AAF22565.1; JOINED.
DR EMBL; AF128416; AAF22565.1; JOINED.
DR EMBL; AF128417; AAF22565.1; JOINED.
DR HSSP; 043570; 1UD0.
DR GeneW; HGNC:1380; CA6.
DR MIM; 144780; -
DR GO; GO:0004089; P:carbonate dehydratase activity; TAS.
DR InterPro; IPR001148; Euk Coahd.
DR Pfam; PF00194; carb_anhydrase; 1.
DR PRODOM; PD000865; Euk Coahd; 1.
DR ProSITE; PS00162; Euk CO2 ANHYDRASE; 1.
KW lysase; zinc; glycoprotein; signal.
FT SIGNAL 1 17
FT CHAIN 18 308
FT METAL 111 111
FT METAL 113 113
FT METAL 138 138
FT METAL 224 224
FT DISULFID 42 42
FT CARBOHYD 67 67
FT CARBOHYD 256 256
FT CONFLICT 90 90
FT CONFLICT 103 103
FT CONFLICT 148 148
FT CONFLICT 270 270
SQ SEQUENCE 308 AA; 35365 MW; 5B6BFF2DD239333 CRC64;

Query Match
Best Local Similarity 20.1%; Score 488; DB 1; Length 308;
Matches 102; Conservative 44; Mismatches 106; Indels 8; Gaps 4;

QY 139 SHWRYG---GDPMPRVSPACAGRFQSPVDIRPOLAFCPALRPELLGQLPPLPELR 194
DB 21 SDWYSGALDEANWPHYFACGQSGSPINLGRTVKRVNPSLKGAMTGYE-TQAGEFP 79
QY 195 LRNGHSVQULTPEGLMALGPGRERVALQLHLHNGAAGR--GSHHTYEGHFRPAELHYH 252
DB 80 WYNNGHVQIGLPEYTEMVADGTVYIAQMFHMGASSEISGSHHTVADGRLHYEIH 139
QY 253 VHLSTARVYDEALGRPGGLAVLAFLBEGPENSAYEQLSLREIEAEGSETOVPLDI 311
DB 140 VHNYSKYTYDIADADAGLAVLAFLBEGVNYENYINSHLANIKYPGQRTTLTGL 199
QY 312 DISALLPSDSRRYFQYSGSLTTPPCAGGVWTFVFNCTWMLSAKQHTLSLDTLMGPDSDL 371
DB 200 DVQMLPEPRNQHVTYTGSLTTPCETENHFWLADFVXLSRTQVKNLSLDRNKTI 259
QY 372 QLNFRATQPLNGRYTASFP 391
DB 260 HNDYRRTQPLKRVVSNFP 279

RESULT 11
CAH6_MOUSE STANDARD; PRT; 317 AA.
ID CAH6_MOUSE STANDARD; PRT; 317 AA.

```

AC P18761; 088625; 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Carbonic anhydrase VI precursor (EC 4.2.1.1) (Carbonate dehydratase  
 DE VI) (CA-VI) (Secreted carbonic anhydrase) (Salivary carbonic  
 DE anhydrase).  
 CN CA6 OR CAR6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Salivary gland;  
 RX MEDLINE=99077987; PubMed=9856573;  
 RA Sok J., Wang X.Z., Batcharova N., Kuroda M., Harding H., Ron D.,  
 RT "Chop-dependent stress-inducible expression of a novel form of  
 RT carbonic anhydrase VI.";  
 RL Mol. Cell. Biol. 19:495-504(1999).  
 RP SEQUENCE OF 18-38.  
 RX MEDLINE=89246331; PubMed=2497732;  
 RA Fenley R.T., Darling P., Aldred P., Wright R.D., Coghlan J.P.,  
 RT "Tissue and species distribution of the secreted carbonic anhydrase  
 RT isoenzyme.";  
 RL Biochem. J. 259:91-96(1989).  
 CC -1- FUNCTION: Reversible hydration of carbon dioxide. Its role in  
 CC saliva is unknown.  
 CC -1- CATALYTIC ACTIVITY: H(2)O(3) = CO(2) + H(2)O.  
 CC -1- COFACTOR: Zinc (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Major constituent of saliva.  
 CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase  
 CC family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF079835; AAD12540.1;  
 DR PIR: S03863; S03863.  
 DR HSSP: P00918; IBV3.  
 DR MGD: MGI:1333786; Car6.  
 DR InterPro: IPR001148; Euk\_Coanhd.  
 DR Pfam: PF00194; carb anhydrase; 1.  
 DR ProDom: PD000865; Euk\_Coanhd; 1.  
 DR PROSITE: PS00162; Euk\_CO2\_ANHYDRASE; FALSE\_NEG.  
 KM Lyase; Zinc; Glycoprotein; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 317 CARBONIC ANHYDRASE VI.  
 FT METAL 110 110 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 112 112 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 137 137 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 41 223 POTENTIAL.  
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 18 18 H -> G (IN REF. 2).  
 FT CONFLICT 35 35 S -> T (IN REF. 2).  
 SQ SEQUENCE 317 AA; 36348 MW; 0393844F7B48283A CRC64;  
 Query Match 19.3%; Score 467.5; DB 1; Length 317;  
 Best Local Similarity 37.4%; Pred. No. 7.4e-20;  
 Matches 98; Conservative 45; Mismatches 106; Indels 13; Gaps 5;

QY 139 SHWRVGD-----PWPVPSVACGRFSGPVDTIPQIAACCPALRPFLIGFPLPPLP 193  
 DB 13 SPMVSYGDDGVQESQMSQYPSCGEERQSPIDVTETVEMFSLKPLSTLVYKENTL-5F 77  
 QY 194 RLRRNGHSVQLTLPGLLMAVLGPGRVYALQLHLHMGAAGR---PQSEHTVGHRRPAAE 249

DB 78 TWNNNGHTVSIIDLPSPMYLETSDGTEP.SKAFFHWG--GRDWELSGSEHTIDGIRIME 135  
 QY 250 IYVHLSSTFAVVDALGRPGGLAVIAA-FLEEGSENSAVQGLSLRLEELAESENOV 308  
 DB 136 AEFVHNKRYGVEYENKQDKNGLAVLAFKIDYAEVYSDIISALKDKERQETTL 195  
 QY 309 PGIDISALLPDSFSEYFQYEGSLTPPCAGVIVTFNQYVLMASKQLTSLDITLWBGD 368  
 DB 196 KDTINDLDPKDVHNYHTYTPGSLTPPTCTENVQVFLDRVTLNQAQVYTIENVMDDNN 255  
 QY 369 SRLQNFRAITOPPLNGVIFASF 390  
 DB 256 NTIQNGYRSTQPNHNVVEANF 277  
 RESULT 12  
 CAHF\_MOUSE STANDARD; PRT; 324 AA.  
 ID CAHF\_MOUSE  
 AC Q99N23;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Carbonic anhydrase XV precursor (EC 4.2.1.1) (Carbonate dehydratase  
 DE XV) (CA-XV).  
 CN CA15 OR CAR15.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Kidney;  
 RA Hewett-Emmett D., Shumlin L.C.,  
 RT "Characterization and evolution of two new members of the alpha-  
 RT carbonic anhydrase gene family in mouse: Car13 and Car15.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC TISSUE=Kidney.  
 CC MEDLINE=22386257; PubMed=12477932;  
 CC Strauberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,  
 CC Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 CC Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 CC Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 CC Datchenko L., Marxson K., Farmer A.A., Rubin G.M., Hong L.,  
 CC Stappleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 CC Brownstein W.J., Ueda T.B., Toshlyuk S., Cantinci P., Prange C.,  
 CC Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 CC Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 CC Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 CC Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 CC Fahy J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,  
 CC Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 CC Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 CC Rodriguez A.C., Grimwood J., Schmutz U., Myers R.W.,  
 CC Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Scherch A., Schen J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Reversible hydration of carbon dioxide (By similarity).  
 CC -1- CATALYTIC ACTIVITY: H(2)O(3) = CO(2) + H(2)O.  
 CC -1- COFACTOR: Zinc (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted. (Potential).  
 CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase  
 CC family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



entires requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL AF231122; AAK16671.1; -

CC EMBL BC019975; AAK19975.1; -

DR HSSP; O64444; 2ZNC.

DR MGD; MGI:1931324; Car15.

DR InterPro; IPR001148; Euk Coand.

DR pfam; PF00194; carb anhydrase; 1.

DR ProDom; PD000865; Euk Coand; 1.

DR PROSITE; PS00162; Euk\_CO2\_ANGHYDRASE; FALSE\_NEG.

KW Lyase; Zinc; Glycoprotein; Signal.

FT CHAIN 1 18

FT METAL 122 122 CARBONIC ANHYDRASE XV.

FT METAL 124 124 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 147 147 ZINC (CATALYTIC) (BY SIMILARITY).

FT CARBOHYD 184 184 ZINC (CATALYTIC) (BY SIMILARITY).

FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 324 AA; 35482 MW; D285DD086476372F CRC64;

Query Match 17.4%; Score 422.5; DB 1; Length 324; Best Local Similarity 35.3%; Pred. No. 2.6e-17; Matches 107; Conservative 42; Mismatches 125; Indels 29; Gaps 8;

QY 136 DDQSHWRVYGGDP-----WPRVSPACAGRFQSPVDIRPOLAFCPALRPLELGFQLP 189

DB 20 DSSGTWCYDSOPKCGPAHWKELAPACGPPQSPINIDLRLVQCDYTLKEFFIQGYDSAP 79

QY 190 LPELRLRNHNSVQTLF-----PGLMALGPREVRALQHLHMGAGRPSEHTVGS 243

DB 80 QDPWLENDGHTVLIRVNSCOQNCFAIRGAGLPSPEYRLLQLHFMWSPGHQSSESLDE 139

QY 244 HRPFAIHYVHLSTAFARVDALGRPGGLAVLAFLSEGPENSAVQLSRLEIAEBS 303

DB 140 KSGMEMHVNHTNYQSWEDARSPDFALVALVLEBDNNTNFSAIVSGKLNSSPG 199

QY 304 SETQVPG-LDISALLPSDS--RYQYEGSLTTPCAQGYWTVPNQTVLISAKQHTLS 360

DB 200 VAVNLTSPFALASLPSALRLRYRYGSLTTPCEBPVALMTVFENTVIRGLAQVQFO 259

QY 361 DTL-WCPGDSR---IQNFRATQPLNGRVIASFPAGVSSPRAEPVQNLSCIAADIL 416

DB 260 AVLGCPGLHRLPLTNSNRPQPLGRRISASPRASVRS-----VSTLPCLH-----L 309

QY 417 ALV 419

DB 310 ALV 312

RESULT 13

CAH2\_TRIHK STANDARD; PRT; 259 AA.

ID CAH2\_TRIHK

AC O8UWAS;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Carbonic anhydrase II (BC 4.2.1.1) (carbonate dehydratase II) (CA-II).

GN CA2.

OS Tribolodon hakonensis (Japanese dace).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Tribolodon.

OC NCBI\_TaxID=151740;

OX NCBI\_TaxID=151740;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2582542; PubMed=12531781;

RA Hirata T., Kaneko T., Ono T., Nakazato T., Furukawa N., Hasegawa S., Wakabayashi S., Shigekawa M., Chang M.H., Romero M.F., Hirose S., RT "Mechanism of acid adaptation of a fish living in a pH 3.5 lake.";

RL Am. J. Physiol. 284:R1199-R1212(2003).

CC -!- FUNCTION: Reversible hydration of carbon dioxide.

CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.

CC -!- COFACTOR: Zinc (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

CC EMBL; AB055617; BAB83090.1; -

DR InterPro; IPR001148; Euk Coand.

DR pfam; PF00194; carb anhydrase; 1.

DR ProDom; PD000865; Euk Coand; 1.

DR PROSITE; PS00162; Euk\_CO2\_ANGHYDRASE; 1.

KW Lyase; Zinc.

FT INIT\_MET 0 0 BY SIMILARITY.

FT METAL 93 93 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 95 95 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 118 118 ZINC (CATALYTIC) (BY SIMILARITY).

SQ SEQUENCE 259 AA; 29002 MW; 622D13D1C2378B86 CRC64;

Query Match 17.4%; Score 422; DB 1; Length 259; Best Local Similarity 36.9%; Pred. No. 2.2e-17; Matches 97; Conservative 46; Mismatches 104; Indels 16; Gaps 8;

QY 139 SH-WRY---GGDPWPRVSPACAGRFQSPVDIRPOLAFCPALRPLELGFQLPPEIR 194

DB 1 SHGNGYADHNGPQKCNENPFIANGPROSPIDITQTKASVYDTIKPKL---KXDPSTSD 57

QY 195 LRNGHSVQTLTPGLEMAL---GP-GREYRALQHLHMGAGRPSEHTVSGHPPAEI 250

DB 58 ILNNGHSFQVTFADNDSSMLTEGPISGKRLQFHHMGASGKSEHTVDSKCPAEI 117

QY 251 HVYHLSAFARVDALGRPGGLAVLAFLSEGPENSAVQLSRLEIAEBSSETQVPG 310

DB 118 HLVENHTKVASFGPAANKPGGLAVGVFDIG-EDNPKQKIDAMDATSKKQKSPFN 176

QY 311 LDISALLPSDSRYFOYEGSLTTPCAQGYWTVPNQTVLISAKQHTLSDTLWCPGDSR 370

DB 177 FDPCLLPKSP-EVYTFGSLTTPFLYESVTWIVCKQPIVSSEQMKKFRSLFLTAEBEK 235

QY 371 ---IQNFRATQPLNGRVIENSP 390

DB 236 ACCWVNNYPPQPLKRVKQCSF 258

RESULT 14

CAH8\_MOUSE STANDARD; PRT; 290 AA.

ID CAH8\_MOUSE

AC P28651; O8CF58; O91XF6;

DT 01-DEC-1992 (Rel. 24, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Carbonic anhydrase-related protein (CARP) (CA-VIII).

GN CA8 OR CAR8 OR CALSI OR CALS OR CARP.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=BALB/c; TISSUE=Brain; MEDLINE=2121526;

RA Kato K.;

RT "Sequence of a novel carbonic anhydrase-related polypeptide and its exclusive presence in Purkinje cells.";

RL FEBS Lett. 271:137-140(1990).

CC [2]

RP SEQUENCE FROM N.A., SEQUENCE OF N-TERMINUS, AND MUTAGENESIS.  
RX MEDLINE=92131620; PubMed=8977131;  
RA Stojedim B., Elledy B., Wallgren K., Jonsson B.-H., Lindskog S.;  
RT Two point mutations convert a catalytically inactive carbonic  
RL anhydrase-related protein (CARP) to an active enzyme.";  
RN FEBS Lett. 398:322-325(1996).  
[3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=Liver;  
RA MEDLINE=22354683; PubMed=1246651;  
RA Okazaki I., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nakada I., Osato N., Saito R., Suzuki H., Yamataka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matzuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brust V., Chochia C., Corbett L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gasteier T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Gimond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kana A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konegaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltas L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan M.J., Petrea G., Pesole G.,  
RA Petrovsky N., Pillai R., Portius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed U.C., Reed D.U., Reid J., Ring B.Z., Ringwald M.,  
RA Sanchelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilmng L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Zimnick P., Hayatsu N.,  
RA Hircano-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Waki J., Waki A., Waki T., Fukuda S.,  
RA Hara A., Hashizume W., Imoto J., Kizawa A., Kizawa T., Kizawa S.,  
RA Yatake A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
[4]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Colon;  
RA MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stachewicz M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshitsuki S., Carninci P., Prange C.,  
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.Y., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.N.,  
RA Butterfield Y.S., Jones S.J.M., Matra M.A.,  
RA Scherch A., Schein J.E., Jones S.J.M., Matra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- FUNCTION: Does not have a carbonic anhydrase catalytic activity.  
CC -1- TISSUE SPECIFICITY: EXPRESSED ONLY IN FURINER CELLS.  
CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase  
CC family.

CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
CC -----  
DR EMBL, X61397; CAA43668.1; -  
DR EMBL, AK004886; BAC25100.1; -  
DR EMBL, BC010773; AAH10773.1; -  
DR PIR, S12867; S12867.  
DR HSP, P00918; IYV3.  
DR MGD, MGI:88253; Car8.  
DR InterPro; IPR001148; Euk Coanhd.  
DR Pfam; PF00194; carb anhydrase; 1.  
DR PROSITE; PS00162; ETK\_CO2\_ANNHYDRASE; 1.  
KW Zinc.  
FT INT MET 0 0  
FT DOMAIN 14 36 GLU-RICH (ACIDIC).  
FT SITE 116 116 ANCESTRAL ZINC LIGAND.  
FT METAL 118 118 ZINC (POTENTIAL).  
FT METAL 141 141 ZINC (POTENTIAL).  
FT MUTAGEN 114 116 EVR->QVR: RESTORE ZINC-BINDING AND  
FT ACTIVITY.  
FT CONFLICT 260 260 A->V (IN REF. 1).  
FT CONFLICT 284 284 V->I (IN REF. 3).  
FT CONFLICT 290 290 Q->S (IN REF. 1).  
SQ SEQUENCE 290 AA; 32950 MW; D0A998A335C8E502 CRC64;  
Query Match 17.3%; Score 419.5; DB 1; Length 290;  
Best Local Similarity 33.3%; Pred No. 3.4e-17; Mismatches 112; Indels 27; Gaps 9;  
Matches 99; Conservative 59; Mismatch 112; Indels 27; Gaps 9;  
QY 112 DLPTVE-APSDPGEPPONNARDEKSDQSHWRGDPMPRVSPACAGRFQSPVDIRPOL 170  
DB 2 DLSTEDVAVAFPEKEDEEEDEEVE---WGVEEGVGLVFPDNGEYQSPINMSRE 58  
QY 171 AAFPPARPELIGFQPP---LPELRNNSHSVQLPLPGLEMAAGP---GEYRAL 223  
DB 59 ARVPS-----LDVRLSPVTVYCRCEVYNDGHTIQLVLSKSVSGGFLPQGGFELY 113  
QY 224 QHLHWGAGRPSSEHTVEGHPPEIHYVHL-STAFARVDEALRGPGIAYIAFLBERG 282  
DB 114 EVRFHWGRENQRGSEHTVNFKAPPELHLIHNSTLFGSIDVAGVPHDIALIFVQIG 173  
QY 283 PEENSAVEQLSLREIEAEGSETQVPGDISALPDSR-YFQYEGSLTPPCAQVI 341  
DB 174 -KEVFGKAVTEILQIQYKSKTICGPNFTLLDPDLRYWYEGSLTTPPSEGT 232  
QY 342 WYFNQTVMSANQL-----HTLSDTLWGPGRSRLQINFRATQPINGRYTEASF 390  
DB 233 WILFRYPLTISQVIEFRRRLRHYVGAELVGCIGLIDNRPQTPLSDRTIRAF 289  
RESULT 15  
CAH8\_HUMAN STANDARD; PRT; 289 AA.  
ID CAH8\_HUMAN  
AC P35219;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Carbonic anhydrase-related protein (CARP) (CA-VIIT).  
DE CA8 OR CALS.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN SEQUENCE FROM N.A.  
RX MEDLINE=93246262; PubMed=8462548;  
RA Skaggs L.A., Bergheim N.C.H., Venta P.J., Tashian R.E.;  
RT "The deduced amino acid sequence of human carbonic anhydrase-related  
RL protein (CARP) is 98% identical to the mouse homologue".  
CC -1- FUNCTION: Does not have a carbonic anhydrase catalytic activity.  
CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase  
CC family.

```
CC -----
CC This SW-SS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L04656; AAA3653.2; -
CC PIR; JN0576; JN0576.
CC HSSP; P00918; 1B73.
CC Genew; HGNC:1382; CAS.
CC MIM; 114815; -
CC GO; GO:0004088; F:carbonate dehydratase activity; TAS.
CC InterPro; IPR001148; Euk_coand.
CC Pfam; PF00194; carb_anhydrase; 1.
CC ProDom; PD000865; Euk_coand; 1.
CC PROSITE; PS00162; Euk_CO2_ANHYDRASE; 1.
CC Zinc.
CC KW INIT MET 0 0 BY SIMILARITY.
CC FT DOMA TN 14 35 GLU-RICH (ACIDIC).
CC FT SITE 115 115 ANCESTRAL ZINC LIGAND.
CC FT METAL 117 117 ZINC (POTENTIAL).
CC FT METAL 140 140 ZINC (POTENTIAL).
CC SQ SEQUENCE 289 AA; 32842 MW; BA6BB5A932A324E CRC64;

Query Match: 17.0%; Score 412; DB 1; Length 289;
Best Local Similarity 33.6%; Pred. No. 9e-17;
Matches 100; Conservative 56; Mismatches 110; Indels 32; Gaps 10;

QY 110 LEDPTVAPGDPQEPQNNARHDKESDDQSHWRYGDDPPWPRVSPACAGRFQSPYDIRQ 163
DB 6 LED--TVAFP--EKEDEEEEEVE--WGIEGVEWGLVFPDANGBYQSPINLSR 56
QY 170 LAAPCAPLPELIGFQLP---LPELRNNGHSYQTLPPGLEMALGP--GREYRA 222
DB 57 EARYDPS-----LDVRLSPNYVYVCRDCEVTNDGHTIQVILKSKSVLSGGPLPQGHFEL 111
QY 223 LQHLHWGAGAPGSGSEHTVEGHRPAEIHVHL-STFAAVDEALGRPGGLAVLAFLBE 281
DB 112 YVRFPHWRENGRSGSEHTVNFKAPMEHLIHMNSTLFGSIDAVGKPHGIAIALFVQI 171
QY 282 GPENSAYEQLLSRLAEIABEGSETOVPGDISALPDSFR-YFOYEGSLTTPCAQGV 340
DB 172 G-KHVGKAVATEIQDYKSKXTPCENPTLPLDPLRDYVWYEGSLTTPCSEGV 230
QY 341 TWTVPNGTVMLSAKL-----HTLSDTLWGPDSRLQANFRATQPLNGRYTEASF 390
DB 231 TWLFRYPPLTISQIEFRRLRTHVKGAEIVGCGDGLGDNFRPTQPLSDRVIRAAF 288
```

Search completed: March 5, 2004, 09:05:30  
Job time: 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 09:00:45 ; Search time 46 Seconds  
(without alignments)  
3148.323 Million cell updates/sec

Title: US-09-967-237-2  
Perfect score: 2424  
Sequence: 1 MAPLCPSMPLILIPAPAPG.....RGTGKGVSPAPVAETGA 459

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mhc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*  
15: SP virus: \*  
16: SP bacteriophage: \*  
17: SP archaea: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	515	21.2	320	6 Q865C0	Q865C0 canis fami
2	493	20.3	344	11 Q8K211	Q8K211 mus musculu
3	465.5	19.2	317	11 Q7TN69	Q7TN69 mus musculu
4	465.5	19.2	325	11 Q80YB7	Q80YB7 mus musculu
5	430.5	17.8	312	13 Q7SYW3	Q7SYW3 xenopus lae
6	409	16.9	264	11 Q811X4	Q811X4 mus musculu
7	397	16.4	260	13 Q7ZK6	Q7ZK6 oncorhynch
8	396.5	16.4	306	13 Q7ZU2	Q7ZU2 brachydano
9	395	16.3	260	6 Q865Y7	Q865Y7 bos taurus
10	380.5	15.7	260	11 Q7TPP1	Q7TPP1 mus musculu
11	379.5	15.7	260	13 Q7ZTU6	Q7ZTU6 xenopus lae
12	379.5	15.6	261	13 Q8UG56	Q8UG56 lepisosteus
13	377.5	15.6	260	13 Q8AVG8	Q8AVG8 xenopus lae
14	375	15.5	243	5 Q8MPH8	Q8MPH8 riftia pach
15	368.5	15.2	251	11 Q811E8	Q811E8 mus musculu
16	360.5	14.9	261	5 Q9XZG6	Q9XZG6 anthopleura

17	354.5	14.6	311	5 Q9VB76	Q9VB76 drosophila
18	349.5	14.4	327	5 Q9W316	Q9W316 drosophila
19	341	14.1	208	4 Q8ETU0	Q8ETU0 homo sapien
20	337.5	13.9	298	5 Q8WRV4	Q8WRV4 aedes aegypt
21	335	13.8	305	13 Q7SZR2	Q7SZR2 brachydano
22	330	13.6	335	5 Q9VTU8	Q9VTU8 drosophila
23	325	13.4	270	5 Q9V396	Q9V396 drosophila
24	325	13.4	717	11 Q8CINO	Q8CINO rattus norv
25	325	13.4	1168	11 Q8CINI	Q8CINI rattus norv
26	325	13.4	1397	11 Q8CIN2	Q8CIN2 rattus norv
27	325	13.4	1426	11 Q8CIN3	Q8CIN3 rattus norv
28	314.5	13.0	1576	13 Q91908	Q91908 xenopus lae
29	314.5	13.0	2271	13 Q91909	Q91909 xenopus lae
30	308.5	12.7	429	11 Q8C4W8	Q8C4W8 mus musculu
31	306.5	12.6	275	16 Q8P3B8	Q8P3B8 xanthomora
32	306	12.6	192	11 Q9DC73	Q9DC73 mus musculu
33	305.5	12.6	1005	4 Q76043	Q76043 homo sapien
34	302	12.5	271	16 Q8PPA4	Q8PPA4 xanthomora
35	301	12.4	306	11 Q83839	Q83839 mus sp. rec
36	298.5	12.3	240	16 Q8D412	Q8D412 vltro vuln
37	297.5	12.3	247	16 Q7VHP3	Q7VHP3 helicobacte
38	297.5	12.3	1612	11 Q9WDH8	Q9WDH8 mus musculu
39	290	12.0	255	2 Q9DXD8	Q9DXD8 rhodopseudo
40	288.5	11.9	319	5 Q02637	Q02637 caenorhabdi
41	286.5	11.8	250	16 Q989X6	Q989X6 rhizobium
42	285.5	11.8	235	5 Q9V9Y8	Q9V9Y8 drosophila
43	281.5	11.6	250	2 Q8KJN6	Q8KJN6 rhizobium
44	281.5	11.6	597	11 Q810U6	Q810U6 mus musculu
45	279	11.5	277	10 Q9FYE3	Q9FYE3 arabidopsis

#### ALIGNMENTS

#### RESULT 1

Q865C0 PRELIMINARY; PRT; 320 AA.  
AC Q865C0;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Carbonic anhydrase VI precursor.  
GN CAC.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_Taxid=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreatic gland;  
RA Murakami M.  
RT "Canine carbonic anhydrase VI (CAC), mRNA."  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB080372; BAC65098.1; -  
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.  
DR InterPro: IPR001148; Euk-Coanhd.  
DR Pfam: PF00194; carb anhydrase; 1.  
DR PRODOM: PD000865; Euk-Coanhd; 1.  
DR PROSITE: PS00162; Euk\_CO2\_ANGHYDRASE; 1.  
KW Signal.  
FT SIGNAL.  
SQ SEQUENCE 1 17 Potential.  
320 AA; 36704 MW; 6E00A8D08A0D23F CRC64;

Query Match 21.2% Score 515; DB 6; Length 320;  
Best Local Similarity 41.6% Pred. No. 2.4e-30;  
Matches 107; Conservative 43; Mismatches 95; Indels 12; Gaps 4;

QY 137 DQSHMYGGDPMPVRVSPACAGRFQSPVDIRPQLAFCAPLRPELLGFGQPPLEIRLR 156  
DB 31 DQVH-----WPREYPTGGTROPSPIDLRKQVNPSTLKKLKGRI-QVGEPFMT 81

```

QY 197 NNGSHVLTLPPLGEMALGPGREYRALQHLHWGAAGR--PGSEHTVEGHPPEAIHVH 254
Db 82 NNGHTVQISLPFTWRMMAVGTEYIAQMHFWHGASSSEISGSEHTIDIRFVAEIHIVH 141
QY 255 LSTAFARVDEALGPGGLAVLAFL-EGPEENSAVEQLSLRLEAEBSGTQVPGDI 313
Db 142 YNSKYKSYDIAQHEPDELAVLALVKEVDYGEHTYNSNISHLNNIRPGSGTVLSGLDI 201
QY 314 SALLPSDFSRFYQYEGSLTTPPAQGVITVFNQVWLSAKQLHTLSDTLMGPGDSRLQL 373
Db 202 EDMLPENTHRYHYRSLTTPCTENWFWLVHVRVLISSIQTMKLNSILDHQQKTLHS 261
QY 374 NFRATQPLNGRVIEASF 390
Db 262 DYRRIQPLNGRVIESNF 278

RESULT 2
ID Q8K2J1 PRELIMINARY; PRT; 344 AA.
AC Q8K2J1;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN 2310047E01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031385; AAH31385.1; -
DR MGD; MGI:1923709; 2310047E01RIK.
DR GO; GO:0004089; F-carbonate dehydratase activity; IEA.
DR GO; GO:0008270; Zinc ion binding; IEA.
DR GO; GO:0006730; Piome-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk_Coand.
DR Pfam; PF00194; carb. anhydrase; 1.
DR ProDom; PD000865; Euk_Coand; 1.
DR PROSITE; PS00162; Euk_CO2_ANHYDRASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 344 AA; 38724 MW; 12DCBD6C3E45D75A CRC64;

Query Match 20.3%; Score 493; DB 11; Length 344;
Best Local Similarity 35.5%; Pred. No. 1,2e-28;
Matches 115; Conservative 50; Mismatches 131; Indels 28; Gaps 7;

```

```

RESULT 3
ID Q7TNG9 PRELIMINARY; PRT; 317 AA.
AC Q7TNG9;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=FVB/N; TISSUE=Salivary gland;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;
RA Klausner R.D.; Collins F.S.; Wagner L.; Shenmen C.M.; Schuler G.D.;
RA Alyschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K.;
RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Heileh F.;
RA Diachenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;
RA Stapleton M.; Soares M.B.; Bonaldo M.F.; Casavant T.L.; Schetz T.E.;
RA Brownstein M.J.; Udell T.B.; Toshnyki S.; Carrinci P.; Prange C.;
RA Raha S.S.; Loggellano N.A.; Peters G.J.; Abramson R.D.; Mullany S.J.;
RA Bosak S.A.; McEwen P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;
RA Richards S.; Worley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Hulyk S.W.;
RA Villalón D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.;
RA Fahey J.; Helton E.; Kettelman M.; Madan A.; Rodriguez S.; Sanchez A.;
RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;
RA Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;
RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield Y.S.;
RA Krzywinski M.I.; Skalska U.; Smalins D.E.; Schnerch A.; Schein J.E.;
RA Jones S.J.; Warr M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N.A.
RA STRAIN=FVB/N; TISSUE=Salivary gland;
RC Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055437; AAH55437.1; -
KW Hypothetical protein.
SQ SEQUENCE 317 AA; 36464 MW; 6A933F95503E58F0 CRC64;

Query Match 19.2%; Score 465.5; DB 11; Length 317;
Best Local Similarity 37.4%; Pred. No. 1,2e-26;
Matches 98; Conservative 44; Mismatches 107; Indels 13; Gaps 5;

```

```

Q80YB7 ID Q80YB7 PRELIMINARY; PRT; 325 AA.
AC Q80YB7;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to carbonic anhydrase 6 (Fragment).
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=salivary gland;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049973.1; AAH49973.1;
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk_COandh.
DR Pfam; PF00194; carb_anhydrase; 1.
DR Prodom; PD000865; Euk_COandh; 1.
FT NON TER 1
SQ SEQUENCE 325 AA; 37235 MW; 58EB4DA1F02FC121 CRC64;

Query Match 19.2%; Score 465.5; DB 11; Length 325;
Best Local Similarity 37.4%; Pred. No. 1.2e-26;
Matches 98; Conservative 44; Mismatches 107; Indels 13; Gaps 5;

QY 139 SHWRVYGD-----PMPRVSPACACRFQSPVDIRPOLAFCAPALRPELLGFCQPPPEL 193
DB 27 SDMSVSGDDGSGESQMSQSEYPCSGERQSPIDVKTIEYMPSLKPSLVNVEKENI-EF 85
QY 194 RLRRNGSHVQLTLPGLLENALGREGYRALQTLHWGAAGR-----PGSEHTVEGHRPPAE 249
DB 86 TWNNQHTVISIDLPKPSMYLETSDGTEFISKAFHFWG--GRDWELSGSEHTIDGIRIME 143
QY 250 IHVYHLSTARFARVDALGRPGGLAVLA--FLBEGSENSAYEQLSLEBLAEAGSETQV 308
DB 144 AHFVFNKRYGYENAKQKQGLAVLAFKIDEVAENTYSDIISALKRIEKGETTTL 203
QY 309 PGLDISALLPDSFSRYFOYEGSLTTPCAQGVITVFNQVTMLSAKQTLHTSLMPCGD 368
DB 204 KDTTIRNLIPKDVHHYTYTPGSLTTPCTENVQFVLRDKVTLKSAQVYVITENSVDHNN 263
QY 369 SRLQINFRATQPLNGRVIEASR 390
DB 264 NTLQNGYRSTQPNRNVVEANF 285

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.D., Ussid T.B., Toshlyuk S., Caxinof P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullany S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Guaranne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boultard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054242; AAH54242.1;
KW Hypothetical protein.
SQ SEQUENCE 312 AA; 34942 MW; DBC9629136044CF1 CRC64;

Query Match 17.8%; Score 430.5; DB 13; Length 312;
Best Local Similarity 35.4%; Pred. No. 4.6e-24;
Matches 99; Conservative 49; Mismatches 107; Indels 25; Gaps 7;

QY 138 QSHWRVYG-----GDPMPRVSPACACRFQSPVDIRPOLAFCAPALRPELLGFCQPP 188
DB 27 EGHWCYELQACDQPCPREFMNNANFPCCGKEKQSPINILTKKAVFWSIKFPLKSYGV- 85
QY 189 PLPELRLRNGSHVQLTLPGLLENALG--GREYRALQTLHWGAAGRPGSEHTVEGHRPP 247
DB 86 SIRQLNTNNGHSAQVLPFGCIBISGGSLGCTDAIQFHHWSEFPQSEHTIDDEKIP 145
QY 248 AEIHVYHLSTARFARVDALGRPGG-----LAVLAFLBEGSENSAYEQLSLEBLAEAG 302
DB 146 MELHIVH-----RRKTKADTGATGSRDLAVLGFFYEHTSTNTDYEPILNSLAIKTK 199
QY 303 GSETQVPGLDISALLP--SDFSRYFOYEGSLTTPCAQGVITVFNQVTMLSAKQTLHTSL 360
DB 200 GAVYKVF-SANITKLIPIKKEELKXYIRNGSLTTPCNEVYTWLTFVTTIKLSQQQURAPY 258
QY 361 DTLNGPGDSRLQINFRATQPLNGRVIEASFPAGVDSPPRA 400
DB 259 NSLFTENQRMVENRFVQRIGDRVYVYISSQAILSSPFS 298

```

```

RESULT 5
Q7SYW3 ID Q7SYW3 PRELIMINARY; PRT; 312 AA.
AC Q7SYW3;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesopatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT Initiative."
RL Dev. Dyn. 225:384-391(2002).

```

```

RESULT 6
Q81IX4 ID Q81IX4 PRELIMINARY; PRT; 264 AA.
AC Q81IX4;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Carbonic anhydrase 7.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen Y., Huang C.-H.;

```

"Molecular identification of carbonic anhydrases (CA) and CA-related (CAR) genes";

RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: A107502.1; AAL78169.1; "

DR GO: GO:0004089; F:carbonate dehydratase activity; IEA.

DR GO: GO:0008270; F:zinc ion binding; IEA.

DR GO: GO:0006730; P:one-carbon compound metabolism; IEA.

DR InterPro: IPR001148; Euk\_Coanhd.

DR Pfam: PF00194; carb anhydrase; 1.

DR Prodom: PD000665; Euk\_Coanhd; 1.

DR PROSITE: PS00162; Euk\_CO2\_ANHYDRASE; 1.

SEQUENCE 264 AA; 29915 MW; B58E0E20CB840FA5 CRC64;

Query Match 16.9%; Score 409; DB 11; Length 264;

Best Local Similarity 37.5%; Pred. No. 1,5e-22;

Matches 101; Conservative 41; Mismatches 99; Indels 28; Gaps 11;

QY 141 WRYG--GDPWPRVSPACAGRFQSPVDIRPOLAFCALPRLLELGFQPLPELRN 197

DB 7 WGYGDDGDSNWKLYPIAQDRSPINISSQAVYSPSLQPLLF--YEACWSLSITN 63

QY 198 NGSVQL-----TLPRG--LEMALPGREYALQHLHWGAAGRPGSEHTVEGHRFP 247

DB 64 NGSVQVDNDSDDRTVVSGLP---GP---YRLKULHFMWKKRMGSSHTVYDGKSPF 117

QY 248 AEIHYVHLST-AFAFVDEALGPGGLAVLAFLLEGPENSAYEQLLSRLEETAEESSEF 306

DB 118 SEIHLVHWAKKYSTFGAALAPDGLAVGVFLETG--DEHSMRLTDLVYVRFKDTKA 176

QY 307 QVPGDLISALPDSRSRYEYEGSLTTPCAQGIYVWYFQNTWLSKQHTLSDTLW-G 365

DB 177 QSCFNPKLPTF-RHYWYVPSLITPLSESTWVILRPIRISQWEKFRSLFTS 235

QY 366 PGDSRLQ--NFRATQPINGRYTEASPPA 392

DB 236 EDDERIHWDNFRPQPLKGRVYKASFOA 264

RESULT 7

Q772K6 PRELIMINARY; PRT; 260 AA.

AC Q772K6; 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Erythrocyte carbonic anhydrase (Salmo gairdneri).

OC Oncorhynchus mykiss (Rainbow Trout) (Salmo gairdneri).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;

OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

NCBI\_TaxID=8022;

SEQUENCE FROM N.A.

RA Esbauch A.J., Lund S.G., Tyffe B.L.;

RT "The Evolution of Red Blood Cell Carbonic Anhydrase: Properties of RBC

CA in a Teleost Fish, Oncorhynchus mykiss."

DL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY307082; AAF73748.1; "

SEQUENCE 260 AA; 28613 MW; SCC928B72B922A52 CRC64;

Query Match 16.4%; Score 397; DB 13; Length 260;

Best Local Similarity 36.1%; Pred. No. 1,1e-21;

Matches 95; Conservative 38; Mismatches 114; Indels 16; Gaps 8;

QY 139 SH-WRYG--GDPWPRVSPACAGRFQSPVDIRPOLAFCALPRLLELGFQPLPELR 194

DB 2 SHWGYGSPDSPEKXCGFPAVANGPRQSPIDITRGQTSYSTLKPKL---KYDQSNSTD 58

QY 195 LANNCHSVQLTPGLMALGPRG---YRALQLHLHWGAAGRPGSEHTVEGHRPAEI 250

DB 59 LANNCHSVQVGVVDVSSSTLTGPTIGIRLKCPRHMGASDSDRSEHTVANGIKFPCEL 118

QY 251 HWVHLSTAFARVDEALGRPGGLAVLAFLLEGPENSAYEQLLSRLEETAEESSEFQVPG 310

DB 119 HLWNNWTKYPSFGGEASEPDGLAVGVFLKIG--AANRLQVLDALCAIKSKQKTFPSN 177

QY 311 LDISALPDSRSRYEYEGSLTTPCAQGIYVWYFQNTWLSKQHTLSDTLW-GPGDS 369

DB 178 FPAKLLLCSD-DWTYDGSITLTPPLESTWVILKSPISVPLQNGKFRSLNFGDGEA 236

QY 370 R--LQNFRAQPINGRYTEASFP 390

DB 237 PCQMDDNYRPPQPLKGRVYKASFP 259

RESULT 8

Q72UE2 PRELIMINARY; PRT; 306 AA.

AC Q72UE2; 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Similar to carbonic anhydrase VII (Fragment).

OC Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

NCBI\_TaxID=7955;

SEQUENCE FROM N.A.

RA Struhsberg R.;

RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC049309; AA049309.1; "

DR GO: GO:0004089; F:carbonate dehydratase activity; IEA.

DR GO: GO:0008270; F:zinc ion binding; IEA.

DR GO: GO:0006730; P:one-carbon compound metabolism; IEA.

DR InterPro: IPR001148; Euk\_Coanhd.

DR Pfam: PF00194; carb anhydrase; 1.

DR Prodom: PD000665; Euk\_Coanhd; 1.

DR PROSITE: PS00162; Euk\_CO2\_ANHYDRASE; 1.

NON TER

SEQUENCE 306 AA; 34587 MW; B35078B0636BC0B CRC64;

Query Match 16.4%; Score 396.5; DB 13; Length 306;

Best Local Similarity 37.3%; Pred. No. 1,5e-21;

Matches 98; Conservative 35; Mismatches 113; Indels 17; Gaps 8;

QY 140 HWRYG--GDPWPRVSPACAGRFQSPVDIRPOLAFCALPRLLELGFQPLPELR 196

DB 48 HMGYGEDNGSAWKKYPIAEGNRQSPIDIVSEAVFPAKSLPIAL---SYNNCTSLIS 104

QY 197 NGSVQLTPPLEMAL--GP--GREYALQLHLHWGAAGRPGSEHTVEGHRPAEIHW 252

DB 105 NGSVSVVEVDTERSVITGGLENNYRLKQPHFMGSKGCCGSEHTVAGKTFVSELHL 164

QY 253 VHL-STAFARVDEALGRPGGLAVLAFLLEGPENSAYEQLLSRLEETAEESSEFQVPG 311

DB 165 VHWNAKXKFSFAAVALPGLAVLGIFLETG--DEHRLHQIDALVYVRFKGSIAEPKGF 223

QY 312 DISALPDSRSRYEYEGSLTTPCAQGIYVWYFQNTWLSKQHTLSDTLWCGCD-- 368

DB 224 NPKCLLPNSL-EWYVPGSLITPLPYESTWVILKEPIYVSEKQKGRFTLLFNGEERD 262

QY 369 -SRQLNFRATQPINGRYTEASFP 390

DB 283 RNEMENNYRPPQPLKGRVYKASFP 305

RESULT 9

Q865Y7 PRELIMINARY; PRT; 260 AA.

AC Q865Y7; 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)



DE Carbonic anhydrase II (EC 4.2.1.1).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 ON NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RA Daigle R., Castro I., Desrochers M., Charest P.-M.,  
 RT "Full length cDNA of Bovine Carbonic Anhydrase II,"  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY240020; AAC85140.1; -  
 DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.  
 DR InterPro; IPR001148; Euk\_Coanhd.  
 DR Pfam; PF00194; carb\_anhydrase; 1.  
 DR ProDom; PD000865; Euk\_Coanhd; 1.  
 DR PROSITE; PS00162; Euk\_CO2\_ANGHYDRASE; 1.  
 KM Lyase.  
 SQ SEQUENCE 260 AA; 29114 MW; A89412C3FFD19X7 CRC64;  
 Query Match 16.3%; Score 395; DB 6; Length 260;  
 Best Local Similarity 35.9%; Pred. No. 1.ee-21;  
 Matches 94; Conservative 42; Mismatches 110; Indels 16; Gaps 8;  
 QY 140 HMYRG---GDPMPRVSPACAGRFQSPVDIRPOLAFCPALRPLELGFQLPPELRIR 196  
 DB 4 HMGYSKNGENHMKDPFIANGRQSPVDIRPOLAFCPALRPLELGFQLPPELRIR 60  
 QY 197 NNGHSVOLTLPPGLEMAL---GP-GREYRALQHLHMGAGRGSEHTVEGHRPAEIHV 252  
 DB 61 NNGHSFVNEVDDSDQKAVLKDGPLTGYRLVGFHFMGSSDDGSEHTVDRKKYAAELHL 120  
 QY 253 VHLSTARVDEALGRGGLAVLAFLBEGPENSAYEQLSRLEIAEGSETQVPGID 312  
 DB 121 VHMNTKYGDPGTAAQDPGLAVGVFLKVG-DANPPLQVLAALSIKTKGKSTDPFNF 179  
 QY 313 ISALLPDSRFYFQEGSLTTPPCAGVITVFNQVTLMSAKOL---HTLSDTLWGPDS 369  
 DB 180 PGLSLPVLND-YMTYPSGLTTPPLESVTWIVAKERISVSSQMLKFTLNNAEGEEL 238  
 QY 370 RLQANFRATOPPLNGRYTEASF 391  
 DB 239 LMLANMRPAQPLKNRQYR-GFP 259

RESULT 10  
 Q7TPE1 PRELIMINARY; PRT; 260 AA.  
 ID Q7TPE1  
 AC Q7TPE1  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Colon;  
 RL MEDLINE=2238257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheet C.F., Bat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Tishiyuki S., Carroll P., Prange C.J.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Maltiani S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettner M., Madan A., Young A.C., Shvechenko Y., Bouffard G.G.,  
 RA Blakesley K.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC055291; AAH55291.1; -  
 DR Hypothetical protein.  
 KM SEQUENCE 260 AA; 29041 MW; 43F81AC12238C5E4 CRC64;  
 Query Match 15.7%; Score 380.5; DB 11; Length 260;  
 Best Local Similarity 35.6%; Pred. No. 1.9e-20;  
 Matches 93; Conservative 41; Mismatches 112; Indels 15; Gaps 8;  
 QY 140 HMYRG---GDPMPRVSPACAGRFQSPVDIRPOLAFCPALRPLELGFQLPPELRIR 196  
 DB 4 HMGYSKNGENHMKDPFIANGRQSPVDIRPOLAFCPALRPLELGFQLPPELRIR 60  
 QY 197 NNGHSVOLTLPPGLEMAL---GP-GREYRALQHLHMGAGRGSEHTVEGHRPAEIHV 252  
 DB 61 NNGHSFVNEVDDSDQKAVLKDGPLTGYRLVGFHFMGSSDDGSEHTVDRKKYAAELHL 120  
 QY 253 VHLSTARVDEALGRGGLAVLAFLBEGPENSAYEQLSRLEIAEGSETQVPGID 312  
 DB 121 VHMNTKYGDPGTAAQDPGLAVGVFLKVG-ASQGLQVLAALSIKTKGKRAAFANFD 179  
 QY 313 ISALLPDSRFYFQEGSLTTPPCAGVITVFNQVTLMSAKOL---HTLSDTLWGPDS 369  
 DB 180 PGLSLPVLND-YMTYPSGLTTPPLESVTWIVAKERISVSSQMLKFTLNNAEGEEL 238  
 QY 370 RLQANFRATOPPLNGRYTEASF 390  
 DB 239 AMYDWMRPAQPLKNRKIKASF 259

RESULT 11  
 Q7ZYU6 PRELIMINARY; PRT; 260 AA.  
 ID Q7ZYU6  
 AC Q7ZYU6  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Similar to carbonic anhydrase II.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 ON NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC041213; AAH41213.1; -  
 DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.  
 DR InterPro; IPR001148; Euk\_Coanhd.  
 DR Pfam; PF00194; carb\_anhydrase; 1.  
 DR ProDom; PD000865; Euk\_Coanhd; 1.  
 SQ SEQUENCE 260 AA; 28806 MW; ACA02D0DD0818F2C CRC64;



RP SEQUENCE FROM N.A.  
 De Gian M.C., Bailly X., Boulben S., Strub J.M., Von Dorsalaer A.,  
 Lallier F.H.;  
 "An insight into molecular and biochemical characteristics of carbonic  
 anhydrases from Riffia pachyptila, a symbiotic invertebrate living  
 under extreme conditions."  
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AJ439711; CAD29128.1; -  
 DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.  
 DR InterPro: IPR001148; Euk Coahnd.  
 DR Pfam: PF00194; carb\_anhydrase; 1.  
 DR Prodom: PD000865; Euk\_Coahnd; 1.  
 DR Prosite: PS00162; Euk\_CO2\_ANHYDRASE; 1.  
 KM Lyase.  
 SQ SEQUENCE 243 AA; 25661 MW; AAA53B407A1P4645 CRC64;

Query Match 15.5%; Score 375; DB 5; Length 243;  
 Best Local Similarity 35.3%; Pred. No. 4.6e-20;  
 Matches 90; Conservative 36; Mismatches 109; Indels 20; Gaps 8;

QY 141 WRVGGDPP--WPRVSPACGRFGQSPVDIRPOLAFCPLRPLELGFOLPPLPELRN 198  
 DB 4 WDYEANGPATWAKSFPLAGKKQSPIDIPASV---SKKSTALVASYNPAASTLTNT 59  
 QY 199 GHSVQLTPGLEMALGP-GREYRALQLHMGAGRPGSEHTVESHRRPADIHVHLST 257  
 DB 60 GLSFQVSVDTGL--SGPPLNGEYKAA5FHFHWSKTSABGESEHTVAGKAYAAEAHIHYNA 117  
 QY 258 A-FARVDEALGRPGGLAVLAFLFEGPEENSAVEQLSLREIEAEESGTQVP-GLDISA 315  
 DB 118 AKYASFDQAVKADGGLAVLAFIIPG-ATNAGVQKIIDLPISVPTGDTATIPGEPDVAC 176  
 QY 316 LLEPDSRIFQYEGSLTTPPCAGVITVENQVMTLSAKQLHTLSDTLWPGDSRIQLNF 375  
 DB 177 LLPGDQSKYWYPPSLTTPPCFBSVTWIVKDIQICENQLALRKI-----TGCNF 228  
 QY 376 RATQPLNGRVTEASF 390  
 DB 229 RPTIGLCGROYSSSF 243

## RESULT 15

Q811E8

ID Q811E8 PRELIMINARY; PRT; 251 AA.

AC Q811E8; 01-JUN-2003 (Tremblrel. 24, Created)  
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Similar to carbonic anhydrase 6 (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=salivary gland;  
 RA Strauberg R.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC046495; AAH46495.1; -  
 DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.  
 DR InterPro: IPR001148; Euk Coahnd.  
 DR Pfam: PF00194; carb\_anhydrase; 1.  
 DR Prodom: PD000865; Euk Coahnd; 1.  
 DR NON TER 1  
 FT NON TER 1  
 SQ SEQUENCE 251 AA; 29182 MW; 17F966486FB4261D CRC64;

Query Match 15.2%; Score 368.5; DB 11; Length 251;  
 Best Local Similarity 38.2%; Pred. No. 1.4e-19;

Matches 78; Conservative 34; Mismatches 85; Indels 7; Gaps 3;  
 QY 192 ELRRNNGHSVQLTPPGLEMALGPGRYRALQLHMGAGR---PGSEHTVESHRRP 247  
 DB 10 EFTWNGHGVISIDLPSSMYLTSDDTEFISKAFFHMG--GRDWELSGSEHTIDGIRSI 67  
 QY 248 AEIHVVALSTAFARVDEALGRPGGLAVLAA-FLFEGPEENSAVEQLSLREIEAEESGT 306  
 DB 68 MEAFVAFNKEVGTENAKQKNGGLAVLAFKIDYAEATYISDLISAKNIEKPGETT 127  
 QY 307 QVPELDISALLPDSRIFQYEGSLTTPPCAGVITVENQVMTLSAKQLHTLSDTLWGP 366  
 DB 128 TLKDTIRNLLPKDVIHYTYPPSLTTPCTENVQFVLADKVTLSKAQVVTIENSVMH 187  
 QY 367 GDSRLQLNFRATQPLNGRVTEASF 390  
 DB 188 NNNTIQNGYRSTQPNHRYVEANF 211

Search completed: March 5, 2004, 09:06:30  
 Job time : 48 secs

RESULT 1  
138013  
p54/58N - human  
C/Species: Homo sapiens (man)  
C/Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 05-Nov-1999  
C/Accession: 138013  
R/Pastorek, U.; Pastorekova, S.; Callebaut, I.; Mornon, J.; Zelnik, V.; Opavsky, R.; Zatl  
Oncogene 9, 2877-2888, 1994  
A/Title: Cloning and characterization of MN, a human tumor-associated protein with a dom  
A/Reference number: 138013; MUID:94366734; PMID:8084592  
A/Accession: 138013  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-459 <RES>  
A/Cross-references: EMBL:X66839; NID:G1000701; P1DN:CAA47315.1; P1D:G1000702  
C/Genetics:  
A/Gene: MatuMN  
C/Superfamily: carbonic anhydrase homology  
F/141-390/Domain: carbonic anhydrase homology <CAH>

Query Match 100.0%; Score 2424; DB 2; Length 459;  
Best Local Similarity 100.0%; Pred. No. 4, 1e-147;  
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAPLCSPWLPPLIPAPAPGLTVQLLSLLLMFVHPQLPRMQEDSFLGGSSGEDDPL	60
DB	1	MAPLCSPWLPPLIPAPAPGLTVQLLSLLLMFVHPQLPRMQEDSFLGGSSGEDDPL	60
QY	61	GEEIDLSEEDSPREDDPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP	120
DB	61	GEEIDLSEEDSPREDDPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLPGEEDLP	120
QY	121	DPEPQNNARHKGDDQSHMRVGGDPMPRVSPACAGRFQSPVDIRPOLAFCPALRPL	180
DB	121	DPEPQNNARHKGDDQSHMRVGGDPMPRVSPACAGRFQSPVDIRPOLAFCPALRPL	180
QY	181	ELIGFQLPPLFELRLRNNGHSHVQLTLPGLMALGPGREYRALQHLHMGAGRGSEHT	240
DB	181	ELIGFQLPPLFELRLRNNGHSHVQLTLPGLMALGPGREYRALQHLHMGAGRGSEHT	240
QY	241	VEGHRPPAEIHVHLSHFARVDENLGRPGGLAVLAFLBEGPEENSAIYQLLSREIEA	300
DB	241	VEGHRPPAEIHVHLSHFARVDENLGRPGGLAVLAFLBEGPEENSAIYQLLSREIEA	300
QY	301	EEGSEIYVGLDISALPLDSFRRYFQYEGSLTTPCAQGVITWTFNOTVMSAKOLHTLS	360
DB	301	EEGSEIYVGLDISALPLDSFRRYFQYEGSLTTPCAQGVITWTFNOTVMSAKOLHTLS	360
QY	361	DTLMGPGSRQLNRAQPLNGRVTEASFPAGTSSPRAAEFVQVNSCLAAGDILALVF	420
DB	361	DTLMGPGSRQLNRAQPLNGRVTEASFPAGTSSPRAAEFVQVNSCLAAGDILALVF	420
QY	421	GLFAVTSVAFLVOMRRORRGTKGVSRPAEVAETGA	459
DB	421	GLFAVTSVAFLVOMRRORRGTKGVSRPAEVAETGA	459